ARMY RESEARCH LABORATORY

The Effect of Variability in Stereo Camera Calibration and Registration Parameters on Three-Dimensional Reconstruction Distance Estimates

By William F. Oberle

ARL-TR-3140 **February 2004**

NOTICES

Disclaimers

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof.

DESTRUCTION NOTICE—Destroy this report when it is no longer needed. Do not return it to the originator.

Army Research Laboratory
Aberdeen Proving Ground, MD 21005-5066

ARL-TR-3140 **February 2004**

The Effect of Variability in Stereo Camera Calibration and Registration Parameters on Three-Dimensional Reconstruction Distance Estimates

William F. Oberle
Weapons and Materials Research Directorate, ARL

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED (From - To)
February 2004	Final	September 2003 to January 2004

4. TITLE AND SUBTITLE

The Effect of Variability in Stereo Camera Calibration and Registration Parameters on Three-Dimensional Reconstruction Distance Estimates

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

William F. Oberle (ARL)

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

U.S. Army Research Laboratory
Weapons and Materials Research Directorate
Aberdeen Proving Ground, MD 21005-5066

8. PERFORMING ORGANIZATION REPORT NUMBER

ARL-TR-3140

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The process of calculating distance from binocular or stereo vision is termed three-dimensional (3-D) reconstruction. Unlike some other approaches for estimating distance, e.g., LADAR (laser detection and ranging), this is not a direct measurement but is a geometric analysis requiring the knowledge of experimentally measured and mathematically derived parameters. Variability in these parameters results in variability in the calculated distances. This report investigates the effect that the variability in these parameters has on computed 3-D reconstruction distances. Results indicate that sub-pixel and milliradian accuracy are required for the parameters in order to have less than a 10% variation in the 3-D reconstruction process.

15. SUBJECT TERMS

camera calibration; 3-D reconstruction; stereo camera registration; stereopsis

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			William F. Oberle
Unclassified	Unclassified	Unclassified	UL	50	19b. TELEPHONE NUMBER (*Include area code*) 410-278-4362

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Contents

List of Figures

List of Tables

Acknowledgments

I would like to thank Thomas Haug of the U.S. Army Research Laboratory (ARL) for his time and effort in reviewing the report. His comments and suggestions proved valuable in the preparation of the final version of the report. Furthermore, I would like to thank Gary Haas, also of ARL, for the many discussions that we had concerning this work.

INTENTIONALLY LEFT BLANK

1. Introduction

To achieve successful autonomous mobility with an unmanned ground vehicle (UGV), the UGV's sensor suite must be capable of accurately determining distances between objects within its environment. This includes not only the distance between itself and objects of interest but also the dimensions (e.g., height) of these objects. A commonly used sensor for this purpose is laser radar[1] (LADAR). LADAR determines distances directly by measuring the time it takes for a laser light pulse emitted by the LADAR sensor to travel from the sensor, reflect off a surface or object, and return to the sensor. Although LADAR provides accurate results, its field of depth is often limited because of the amount of energy that can be imparted to the laser pulse. Other negatives associated with LADAR include the high cost of high resolution units and the active nature of the sensor that may not be acceptable in all tactical situations.

An alternate approach to LADAR for distance measurement is stereopsis[2]. In this approach, images from two cameras[3] of the same scene taken simultaneously are analyzed to determine distances. Since cameras rely on incident light to generate images, this approach is passive in nature, and generally, the distances at which objects can be identified are greater than with LADAR. In addition, high quality, high resolution cameras, e.g., those capable of producing images with 1000+ pixels in width by 1000+ pixels in height, are readily available for thousands of dollars. LADAR units, on the other hand, with image resolution on the order of 180 pixels by 32 pixels, cost tens of thousands of dollars. The major drawback of the stereopsis approach is that distance is not determined directly as in LADAR but is calculated with a geometric analysis termed three-dimensional (3-D) reconstruction (see Oberle and Haas, 2002, or Trucco and Verri, 1998 for a detailed description). As described in these references, the 3-D reconstruction analysis employs experimentally determined parameters for each camera as well as the stereo images recorded by the stereo camera pair. The process of obtaining the necessary parameters for a camera is termed "camera calibration" or simply calibration and involves the analysis of a set of calibration images of a special calibration poster taken by the camera being calibrated. Thus, the accuracy of distance calculations based on stereopsis depends on the accuracy of the camera calibration.

In theory, if the camera settings (e.g., focal length) are not changed, the camera parameters should remain fixed. Unfortunately, in practice, this has not proved to be the case. Pre- and post-calibration results for cameras used to experimentally collect data in off-road environments have shown differences. Additionally, when we have performed the calibration analysis using

[1] Also referred to as laser detection and ranging.

[2] The determination of depth or distance attributable to binocular vision.

[3] The two cameras are referred to as the stereo camera pair. In some implementations, more than two cameras are used. However, two cameras are sufficient to determine distance.

different calibration image sets, even if they were obtained at approximately the same time (within minutes) and without moving or changing the settings on the camera, different results for the camera parameters have been obtained. This implies that in the 3-D reconstruction analysis, the camera parameters will only be approximately known. What impact the uncertainty in the camera parameters has on the 3-D reconstruction is the focus in this report. Specifically, we will empirically assess the sensitivity of the 3-D reconstruction results (i.e., distance calculation) attributable to uncertainty or variability in the computed camera parameters.

The remainder of the report is organized as follows. We begin with a discussion of the experimentally obtained and derived camera parameters, studying a specific set of calibrations to illustrate the variability in the calibration results. Next, we empirically assess the impact that this variability can have on the 3-D reconstruction results, first, in section 3, for feature-based 3-D reconstruction, followed by area-based 3-D reconstruction using rectified images and dense stereo matching in section 4. Distance or range resolution and its impact on the 3-D reconstruction process is discussed in the fifth section. The final section includes a summary of the work and recommendations.

2. Camera Calibration and Parameter Variability

Three sets of parameters must be determined in order for us to perform the 3-D reconstruction. Two of the three, the *intrinsic* and *extrinsic* camera parameters, are experimentally determined through the calibration process. The third set of parameters is derived from the first two.

Ignoring lens distortion and assuming a perspective camera model, the *intrinsic* parameters (Trucco and Verri, 1998) are the focal length, f, in millimeters, the location of the image center in pixel coordinates, (o_x,o_y), and the effective pixel size (millimeters) in the horizontal and vertical direction, (s_x,s_y). These parameters link the pixel coordinate of an image point with the corresponding coordinates in the camera coordinate system. Individual knowledge of f, s_x, and s_y is not required to perform the 3-D reconstruction; it is sufficient to know the ratios f/s_x and f/s_y, termed the horizontal and vertical pixel pitch in this report. Generally, it is these ratios that are experimentally estimated during the calibration process. These ratios are also interpreted as the camera focal length measured in pixels. The *extrinsic* camera parameters for each camera specify the transformation between the camera and the world coordinate system (Trucco and Verri, 1998). For the purposes of 3-D reconstruction, however, it is the transformation between the two camera coordinate systems that is required. This transformation is obtained from the intrinsic and extrinsic camera parameters through straightforward matrix/vector calculations (Trucco and Verri, 1998). It is this derived set of parameters that we are concerned with in this report. We refer to these parameters as *registration* parameters. Specifically, the *registration* parameters are the translation vector, **T**, and the rotation matrix, **R**, which specify the

transformation between the coordinate system of the left camera of the stereo camera pair to the coordinate system of the right camera of the stereo camera pair. Symbolically, if p_l represents the coordinates of a point in the coordinate system of the left camera and p_r the coordinates of the same point in the coordinate system of the right camera, the *registration* parameters satisfy the equation, $p_r = \mathbf{R}(p_l - \mathbf{T})$.

Although a detailed description of the calibration process is outside the scope of this report, an abbreviated description is provided to aid in understanding potential sources of error in the *intrinsic* and *extrinsic* parameters. A more detailed description of the calibration process that we use is contained in an earlier report (Oberle and Haas, 2002). The calibration process in general is described in many texts on computer vision, for example, Trucco and Verri (1998) or Faugeras (1993).

Our calibration process is based on a facility design[4] and software (Litwin, 2000) provided by the Jet Propulsion Laboratory (JPL), California Institute of Technology, Pasadena, California, in combination with software developed in house (Oberle and Haas, 2002). The calibration for each camera of the stereo camera pair requires four images of a calibration poster (calibration image set) at four surveyed stations (left rear, right rear, right front, and left front). An example of the required calibration images is shown in figure 1. The JPL software processes the calibration images together with the surveyed geometry of the calibration poster and stations to yield a camera model in the ***CAHVOR***[5] format as described in Yakimovsky and Cunningham (1978), Gennery (2001), and JPL (2002). The ***CAHVOR*** model directly provides the values of the *intrinsic* parameters. Results of the ***CAHVOR*** model are used as input to our in-house software to compute the *extrinsic* parameters of the camera. Once the *extrinsic* parameters for the left and right cameras of the stereo camera pair are computed, the *registration* parameters for the stereo camera pair are determined.

The experimental data used in this report were provided by Gary Haas[6] and consist of five calibration image sets used in the calculation of the camera parameters and three stereo image pairs used in the 3-D reconstruction. The calibration image sets were recorded in May (pre) and September (post) 2003. To our knowledge, the camera settings (i.e., focal length) and their relative location with respect to each other[7] remained unchanged between May and September. The stereo camera system was used to obtain stereo images at the U.S. Army Research Laboratory (ARL) and National Institute of Standards and Technology from 8 May 2003 through 20 May 2003.

---

[4]Private communication, Larry Matthies, JPL, California Institute of Technology, Pasadena, California, 1997.

[5]not an acronym

[6]Private communication, Gary Haas, ARL, Aberdeen Proving Ground (APG), Maryland, September 2003.

[7]The cameras remained undisturbed and mounted in the rigid frame of the stereo "rig" during the time frame over which the image sets were taken.

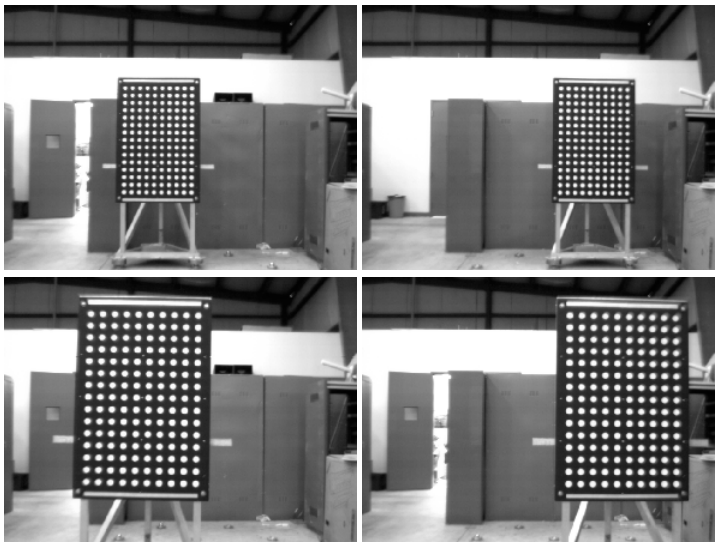

Figure 1. Images of the calibration poster in the four required locations, clockwise, left rear, right rear, right front, and left front (Oberle and Haas, 2002).

The calibration image sets are labeled 1 through 5. Table 1 provides the dates on which the calibration image sets were acquired. Thus, calibration image set 1 was obtained before the data acquisition in the field while calibration image sets 2 through 5 were obtained afterward.

Table 1. Information about the image sets used in the calibrations.

Image Set Label	Date of Images
1	8 May 2003
2	16 September 2003
3	17 September 2003
4	17 September 2003
5	17 September 2003

The three calibration image sets obtained on 17 September were taken at approximately the same time without changes in the camera settings. Camera settings were not changed between 16 September and 17 September either. Details about the three stereo image pairs used in the 3-D reconstruction are provided in the next section.

Results for the left and right camera *intrinsic* parameters for the five image sets are given in tables 2 and 3. Calibration images are 640 pixels wide (horizontal) and 480 pixels high (vertical). In addition to the computed *intrinsic* parameter values, each column contains the average, standard deviation, range, and coefficient of variation (CoV) for the data in the column. We use the CoV, defined as the absolute value of the ratio of the standard deviation to the average expressed as a percentage, as our principal measure of variability.

Table 2. Image center intrinsic parameter results for the different image set calibrations.

Image Center Location (pixels)				
	Left Camera		Right Camera	
Image Set Label	Horizontal (x)	Vertical (y)	Horizontal (x)	Vertical (y)
1	334.69	250.46	344.19	264.43
2	336.77	246.27	346.65	265.40
3	335.05	247.87	347.92	263.35
4	337.15	248.94	350.16	267.30
5	335.25	248.56	349.16	266.69
Average	335.78/336.06	248.42/247.91	347.62/348.47	265.43/265.69
SD*	1.10/1.06	1.53/1.18	2.32/1.52	1.61/1.75
Range	2.46/2.1	4.19/2.67	5.97/3.51	3.34/3.34
CoV	0.33%/0.32%	0.62%/0.48%	0.67%/0.44%	0.61%/0.66%

*SD = standard deviation

Table 3. Pixel pitch intrinsic parameter results for the different image set calibrations.

Pixel Pitch (focal length/pixel length)				
	Left Camera		Right Camera	
Image Set Label	Horizontal (f/sx)	Vertical (f/sy)	Horizontal (f/sx)	Vertical (f/sy)
1	866.02	866.00	852.59	852.34
2	865.89	865.57	852.21	851.72
3	865.15	865.03	852.92	852.56
4	865.63	865.35	852.27	851.63
5	865.20	864.43	852.74	851.83
Average	865.58/865.47	865.28/865.10	852.55/852.54	852.02/851.94
SD	0.39/0.35	0.59/0.50	0.30/0.35	0.41/0.42
Range	0.87/0.74	1.57/1.14	0.71/0.71	0.84/0.84
CoV	0.045%/0.040%	0.068%/0.058%	0.035%/0.041%	0.048%/0.049%

Since calibration image set 1 was obtained before the data acquisition in the field, two values for average, standard deviation, range, and CoV are provided. The first value corresponds to using the results for the five calibration image sets, whereas the second value is the results with the calibration image sets acquired at approximately the same time (i.e., image sets 2 through 5).

Referring to table 2, the variability in computed image center location is evident. However, the pixel pitch results in table 3 indicate a much lower degree of variability. Since it is unlikely that the effective pixel size, (s_x,s_y), could change, the results in table 3 strongly suggest that the focal length settings did not significantly vary, supporting our belief that the camera settings remained fixed during the time frame of the experimental work.

As discussed in our earlier report (Oberle and Haas, 2002), the JPL calibration software requires the user to mark the centers of the four corner white circles on the calibration poster (see figure 1). Subsequently, the program estimates the location of the centers of all 150 white circles on the calibration poster. Thus, the JPL calibration calculation is not completely deterministic, containing at least the potential for a degree of randomness or noise in the selected location of the centers of the white circles in the image(s) of the calibration poster. Is it possible that this

randomness produces the variation observed in the calculation of the image center location observed in table 2? To address this question, the calibration calculation is repeated on a single calibration image set. In this case, calibration image set 2 is selected. Results for the center location are provided in table 4.

Table 4. Image center intrinsic parameter results for repeated calibration calculations with image set 2.

Center Location (pixels) Repeated Calibration With Image Set 2				
	Left Camera		**Right Camera**	
Calibration	**Horizontal (x)**	**Vertical (y)**	**Horizontal (x)**	**Vertical (y)**
1	336.77	246.27	346.65	265.40
2	336.80	246.29	346.64	265.42
3	336.79	246.30	346.63	265.44
4	336.78	246.29	346.65	265.41
5	336.78	246.28	346.62	265.45
Average	336.78	246.29	346.64	265.42
SD	0.01	0.01	0.01	0.02
Range	0.02	0.03	0.03	0.05
CoV	0.003%	0.004%	0.003%	0.008%

A comparison of tables 2 and 4 shows a decrease by roughly two orders of magnitude in the CoV. This would seem to eliminate any potential randomness in the JPL calibration software as being the source of the variation in the calculation of the image center location.

Next, we address the *registration* parameters. Tables 5 through 8 provide the results for the *registration* parameters from the calibration calculations using the five image sets. Results for the translation vector are presented in table 5. We provide only the average rotation matrix (table 6), i.e., each entry in the matrix is the average of the corresponding entries in the five calculated rotation matrices, the standard deviation for each rotation matrix entry (table 7), and the CoV for each rotation matrix entry (table 8).

Table 5. Translation vector extrinsic parameter results for the different image set calibrations.

Translation Vector (mm)			
Image Set Label	**Horizontal (x)**	**Vertical (+y down)**	**Forward (z)**
1	334.616	-1.479	-3.019
2	334.700	-1.734	-2.476
3	335.007	-1.830	-5.372
4	334.749	-1.564	-1.928
5	334.734	-1.771	-6.839
Average	334.761/334.798	-1.676/-1.725	-3.927/-4.154
SD	0.146/0.141	0.148/0.114	2.091/2.343
Range	0.391/0.307	0.351/0.266	4.911/4.911
CoV	0.044%/0.042%	8.831%/6.609%	53.247%/56.403%

Table 6. Average rotation matrix extrinsic parameter results for the different image set calibrations.

Average Rotation Matrix Image Sets 1 – 5		
0.99980080/0.99979950	0.01648904/0.01639748	-0.01046820/-0.01072768
-0.01630862/-0.01623388	0.99986580/0.99986675	-0.00382917/-0.00384633
0.01041726/0.01068068	0.00399478/0.00401568	0.99993600/0.99993275

Table 7. Standard deviation for each entry in the rotation matrix extrinsic parameter results for the different image set calibrations.

Standard Deviation for Each Cell Average Rotation Matrix Image Sets 1 – 5		
0.00001486/0.00001682	0.00024763/0.00016083	0.00165424/0.00178880
0.00027697/0.00025502	0.00000396/0.00000386	0.00134050/0.00154724
0.00166038/0.00179255	0.00134217/0.00154886	0.00001844/0.00001957

Table 8. Coefficient of variation for each entry in the rotation matrix extrinsic parameter results for the different image set calibrations.

Coefficient of Variation for Each Cell Average Rotation Matrix Image Sets 1 – 5		
0.0001%/0.0017%	1.5%/0.9808%	15.8%/16.675%
1.7%/1.571%	0.0004%/0.0004%	35.0%/40.226%
15.9%/16.783%	33.6%/38.570%	0.0018%/0.0020%

First, the translation vector is discussed. From table 5, the horizontal component exhibits little variability, i.e., CoV = 0.044%. On the other hand, the vertical and especially the forward components vary substantially with CoVs equal to 8.831% - 6.609% and 53.34% - 56.403%, respectively.

As for variation in the rotation matrix, the results in table 8 for the CoV for each matrix entry indicate that there is substantial variability in the non-diagonal components of the rotation matrix.

An alternate approach to quantify the variability in the *registration* parameters is not to analyze individual components of the *registration* parameters but to combine the components into a single measure for each parameter. For the translation vector, we chose the angle between two translation vectors obtained using the scalar vector product. Results are shown in table 9. The large CoV for the vertical and forward components of the translation vectors in table 5 result in what we consider large angles (> 3 milliradians) between the vectors. Since the vector scalar product is commutative, only the upper triangular portion of the table needs to be computed. Besides indicating overall variability, these results can also be used to investigate the variability associated with the translation vector for each calibration image set by computing the average between the translation vector for one of the calibration image sets and the other four calibration

image sets. This average is given in the last line of the table. With this metric, the translation vector for calibration image set 1 is the best fit to the other translation vectors with an average angle of 5.89 ([1.79 + 7.09 + 3.27 + 11.40]/4) milliradians between the other translation vectors.

Table 9. Angle between calibration image set translations.

Angle Between Translation Vectors (milliradians)					
Image Set Label	**1**	**2**	**3**	**4**	**5**
1		1.79	7.09	3.27	11.40
2			8.64	1.72	13.03
3				10.30	4.40
4					14.68
Average Angle	5.89	6.30	7.61	7.49	10.88

For the rotation matrix of the *registration* parameters, besides an examination of the behavior of each component, two other approaches to quantify variability are provided. First, a single angle is chosen to represent the variability. We obtain the angle by employing quaternions[8]. If A and B represent the matrices associated with two rotations between the left and right camera coordinate systems for two calibration image sets, then $B^{-1}A$ represents a rotation that rotates the left camera coordinate system to the right camera coordinate system and back. The closer the two rotations are to each other, the closer this product, $B^{-1}A$, is to the identity matrix. Now any rotation can be represented as a rotation of an angular measure about a fixed axis. The angle of rotation is the scalar component of the unit quaternion associated with the rotation. Using the angle from the quaternion does not completely describe the rotation since we are not providing the vector about which the rotation is performed. However, we want to provide a single measure and feel that the amount of angular rotation provides more information. Results are provided in table 10. The average of the rotation angle from all calculations involving a particular rotation matrix for one of the calibration image sets is shown in the diagonal element of the table. As with the translation vectors, it appears that the rotation matrix associated with calibration image set 1 is the best fit to the other rotation matrices with an average angular rotation of 2.865 ([0.844 + 3.490 + 2.670 + 3.911 + 2.086 + 3.232 + 2.706 + 3.981]/8) milliradians. As mentioned before, the direction about which the rotation is performed is not included in the rotation angle.

The second approach involves a more complete description of the variability in the rotation matrix. This is provided through the use of roll, yaw, and pitch angles. Roll is a rotation about the forward axis, yaw a rotation about the vertical axis, and pitch a rotation about the horizontal axis. Assuming that the order in which the roll, yaw, and pitch rotations are applied is fixed, then there is a one-to-one correspondence between these three angles and the rotation matrix

[8]Private communication, G. Haas, ARL, APG, MD, November 2003.

(Oberle, 2003). We use the order pitch followed by yaw followed by roll. Results for the roll, yaw, and pitch angles for the five different calibration image set rotations are given in table 11.

Table 10. Rotation angle associated with the rotation obtained by rotating from the left camera coordinate system to the right camera coordinate system and back for the rotation matrices from the calibration image sets (average rotation angle for rotation i located in i^{th} diagonal element).

Rotation Angle of $B^{-1}A$ (milliradians)						
		Inverse Matrix (B^{-1})				
	Image Set Label	**1**	**2**	**3**	**4**	**5**
A	**1**	2.865	0.844	3.490	2.670	3.911
	2	2.086	3.790	4.905	4.043	5.409
	3	3.232	4.323	4.067	4.021	4.018
	4	2.706	3.591	4.254	3.054	1.515
	5	3.981	5.116	4.295	1.630	3.734

Table 11. Roll, yaw, and pitch angles for the rotations of the calibration image sets.

Roll, Yaw, Pitch (milliradians)			
Image Set Label	**Roll**	**Yaw**	**Pitch**
1	-16.60909172	-9.363706833	3.912291491
2	-16.10733124	-8.14870018	4.179730939
3	-16.58617967	-11.08622709	1.780300343
4	-16.22340857	-11.1108286	5.167850574
5	-16.01521139	-12.37781607	4.935818144
Average	-16.308/-16.233	-10.417/-10.681	3.995/4.016
SD	0.2744/0.2504	1.6604/1.7926	1.3423/1.5490
CoV	1.683%/1.543%	15.939%/16.783%	33.599%/38.571%

As with the image center location calculation, the possibility that the variability observed in the *registration* parameters is attributable to the randomness described earlier in the JPL calibration software must be addressed. Table 12 presents the results for the translation vector obtained from the repeated calibration calculation on calibration image set 2 used earlier in the discussion of the image center locations.

A comparison of tables 5 and 12 shows a decrease in the CoV of approximately an order of magnitude for each component. However, the CoV for the forward component remains large. For the rotation matrix, we again compute the average matrix from the five repeated calibrations of image set 2 and determine the CoV for each of the entries. The results for the CoV are presented in table 13 and indicate a substantial decrease in the CoV for the entries in the average rotation matrix.

Table 12. Translation vector extrinsic parameter results for the repeated calibration calculation of image set 2.

Translation Vector (mm) Repeated Calibration With Image Set 2			
Calibration	**Horizontal (x)**	**Vertical (+y down)**	**Forward (z)**
1	334.702	-1.810	-3.043
2	334.678	-1.804	-2.958
3	334.694	-1.795	-2.960
4	334.701	-1.781	-3.112
5	334.688	-1.788	-2.877
Average	334.693	-1.795	-2.990
SD	0.010	0.012	0.090
Range	0.024	0.019	0.235
CoV	0.003%	0.67%	3.01%

Table 13. CoV results for the average rotation matrix for repeated calibration calculation of image set 2.

Coefficient of Variation for Each Cell Average Rotation Matrix Repeated Calibration With Image Set 2		
0.00004%	0.03157%	0.19584%
0.13388%	0.00053%	0.33503%
0.20152%	0.32663%	0.00004%

At least for the set of data analyzed, it appears that the camera calibration results in a degree of variability in the calculated *intrinsic*, *extrinsic*, and *registration* parameters. This variability appears to impact all parameters with the exception of pixel pitch. In general, it does not appear that this variability can be attributed to the randomness associated with the actual calibration calculation. Since the principal use of these parameters is 3-D reconstruction, it remains to be determined to what extent this variability impacts the results of the 3-D reconstruction. This topic is explored in the next two sections.

3. Feature-Based 3-D Reconstruction

In feature-based 3-D reconstruction, pixels corresponding to the same feature (e.g., a corner) in the left and right images of the stereo image pair are identified. Generally, this results in a sparse set (less than 1% of the total number of pixels) of matching points. Once the matching points are identified, 3-D reconstruction by triangulation is performed to determine distance (see figure 2), since base angles and length of base are known. However, since the camera parameters are only approximately known (as indicated by the variability in the calibration calculation results), even if the matching point locations are exact, the two rays will not necessarily intersect in space. In

this case, their intersection is approximated as the point of minimum distance to both rays (Oberle and Haas, 2002; Trucco and Verri, 1998). This case is illustrated in figure 3.

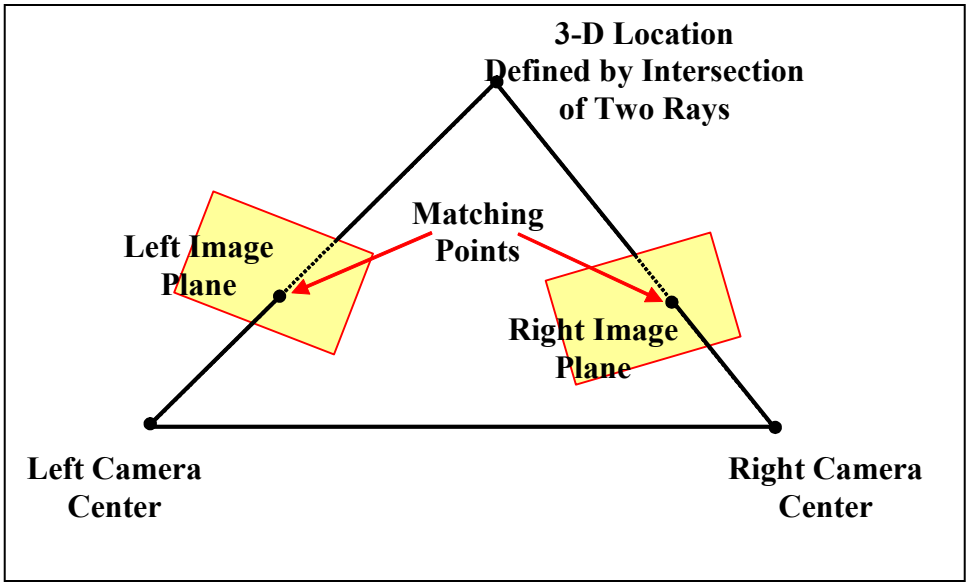

Figure 2. Triangulation with intersecting rays.

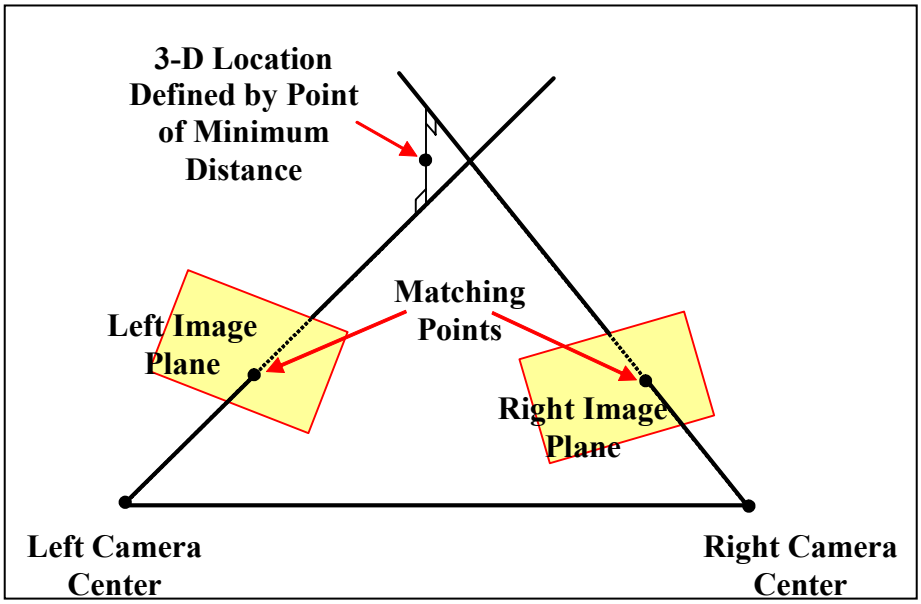

Figure 3. Triangulation when rays do not intersect (adapted from Trucco and Verri, 1998).

As can be seen in figure 3, small differences in the location of the rays can result in a substantial change in the resulting 3-D location. The situation will be exacerbated with distance from the camera centers.

Three stereo image pairs are used in the feature-based 3-D reconstruction analysis. The stereo image pairs are shown in figures 4 through 6 with the left camera image on the left and the right

camera image on the right. To distinguish the image pairs, the following nomenclature is used. Figure 4 is the building image pair, figure 5 is the intersection image pair, and figure 6 is the field image pair. The building image pair was obtained on 8 May 2003, while the other two image pairs were obtained on 20 May 2003.

Figure 4. Building image pair.

Figure 5. Intersection image pair.

Figure 6. Field image pair.

We selected corners as the image feature for the feature-based 3-D reconstruction. To determine corresponding or matching corner points in the left and right images, a two-step procedure is used. First, a Harris corner detector (Harris and Stephens, 1988) implemented by Torr (2002) in MATLAB[9] (The MathWorks, 2001) is used to identify pixel locations in the left and right images that represent corners. This produces two sets of corner points, one for the left image, C_L, and one for the right image, C_R. In step two, each corner point in C_L is paired with its best match in C_R. A 7-by-7 correlation window is used. The similarity measure used in the correlation is the sum of squared differences. This process does not guarantee that all matches are correct. Thus, the 3-D reconstructed results will most likely contain outliers. One set of outliers is those matched points for which a negative forward value is computed in the 3-D reconstruction since all scene points are situated in front of the cameras (positive forward value). These points are deleted from the 3-D reconstruction results.

3.1 Results for Building Image Pair

For the building image pair, 272 matched corner points were identified. The corner points ascertained by the Harris corner detector for each image are shown in figure 7. Note that most of the identified points are situated at approximately the same distance from the stereo camera pair in the trees behind the buildings. From direct measurement, the tree line is approximately 50 meters' distance from the cameras. Using the methodology discussed in Trucco and Verri (1998) implemented in our in-house software (Oberle and Haas, 2002), we performed the 3-D reconstruction for these points for each set of camera parameters presented in the previous section. Of the 272 matched points, 265 points are considered valid (i.e., 7 points are outliers) based upon the 3-D reconstruction. Results are provided in figure 8.

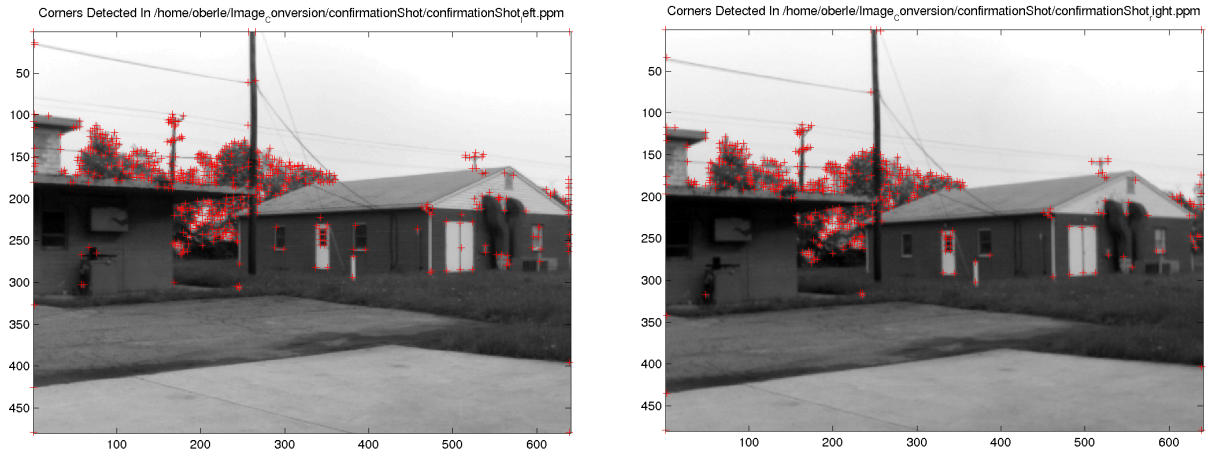

Figure 7. Corner points (+) as determined by the Harris corner detector for the building image pair (left image on the left and right image on the right).

[9]MATLAB® is a registered trademark of The MathWorks.

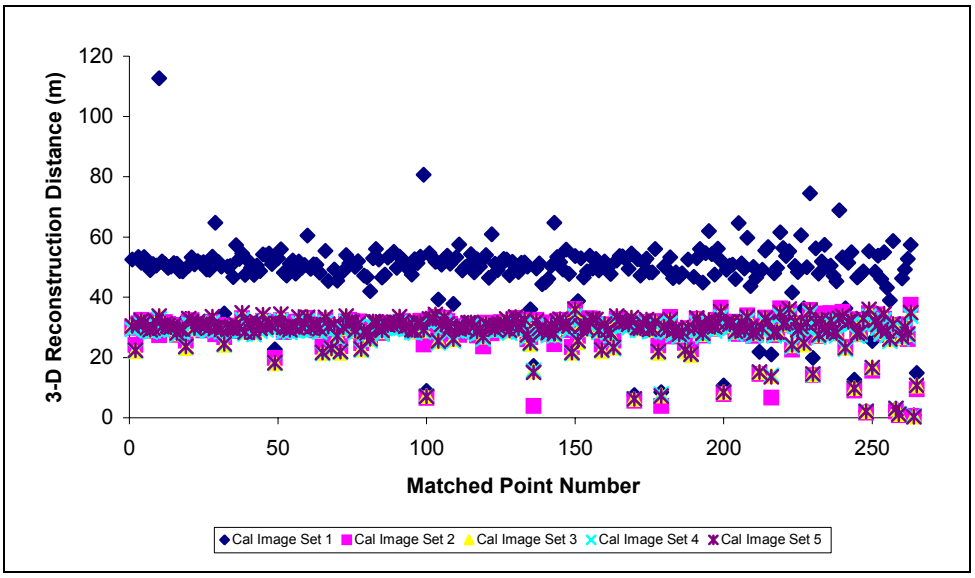

Figure 8. 3-D reconstruction results for the building image pair with the five different sets of camera parameters.

The horizontal axis corresponds to the 265 valid matched points, and the vertical axis is the 3-D reconstruction distance with the different calibration results. The order on the horizontal axis is arbitrary, based solely on the order in which corner points in the left image were identified. To interpret the information in the graph, one must inspect the five different 3-D reconstruction distances corresponding to the five calibrations along a given vertical line. This is illustrated in figure 9 for several of the points in the matched point range of 50 to 70 from figure 8. Ideally, one would like the five distance estimates to coincide.

Returning to figure 8, it appears that two distinct solutions are present, one corresponding to calibration image set 1 and a second for the other four calibration image sets. Based on the measured distance, the results for calibration image set 1 appear to more accurately reflect the actual distances. From the previous section, it appeared that the *registration* parameters of calibration image set 1 were the best fit for the five sets of parameters. Thus, at least for this image pair and for the distances computed, it would appear that the differences in the *intrinsic* parameters, specifically the image center location, have the greatest impact on the 3-D reconstruction.

The objective of this report is not to determine which of the calibrations produces the best results but to investigate the variability in the 3-D reconstruction distance estimates as a function of the variability in the calibration values. To investigate this question, the spread (range in statistical terminology, i.e., maximum value minus minimum value) of the calculated 3-D reconstruction distance is analyzed. Instead of plotting the spread versus matched point number, the spread is plotted versus the average calculated 3-D reconstruction distance. Specifically, for each matched point, the five different calculated 3-D reconstruction distances corresponding to the five different calibration image sets are averaged; this is the abscissa. The corresponding ordinate is

the spread or range between the five distance estimates. We use the term "spread" in place of "range" to avoid confusion since the term range is often used in conjunction with discussions of distance and location. In figure 10, results using all five calibration image sets are presented. Results using calibration image sets 2 through 5 are shown in figure 12. As would be expected given the results shown in figure 8, there is substantial spread (i.e., large value for the range) in the calculated 3-D reconstruction distances for each pair of matched points as shown in figure 10. In addition, the magnitude in the spread increases as the average 3-D reconstruction distance increases and the relationship appears to be slightly nonlinear. The spread or variability in the resulting 3-D reconstruction distances would appear to be too large to provide useful distance information.

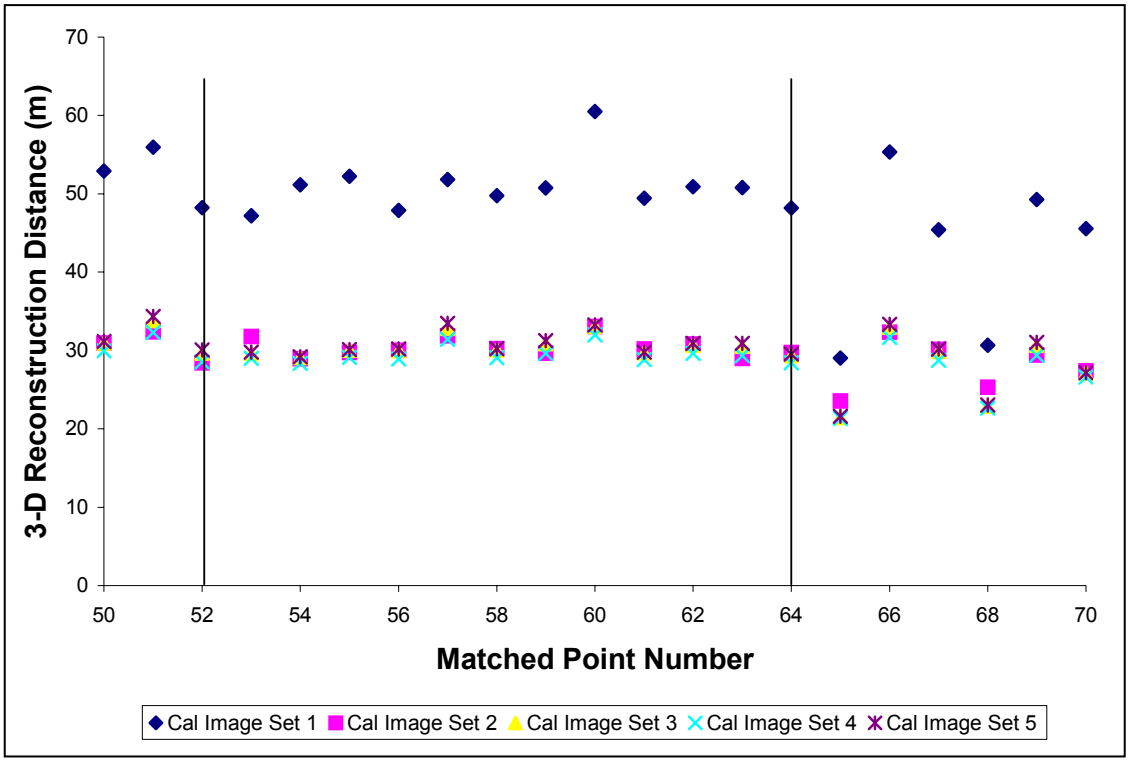

Figure 9. Example of how to interpret 3-D reconstruction results for selected points from figure 8.

To quantify this statement, we determine the percent that the spread represents of the average 3-D reconstruction distance. Essentially, this value indicates how precisely the 3-D reconstruction distance is calculated for the given variability in the camera parameters at the given distance. For example, if at an average distance of 30 meters, the spread is 20%, then the actual distance should be 30 ± 6 meters (20% of 30 = 6). The best fit to the data in figure 10 is $y = 0.3261x^{1.1779}$. Dividing by the average distance, x, and multiplying by 100, the percentage of the spread relative to the average 3-D reconstruction distance is given by $y = 32.61x^{0.1779}$. Results are shown in figure 11, indicating that the spread represents over 35% of the average 3-D reconstruction distance for distances more than about 1 meter.

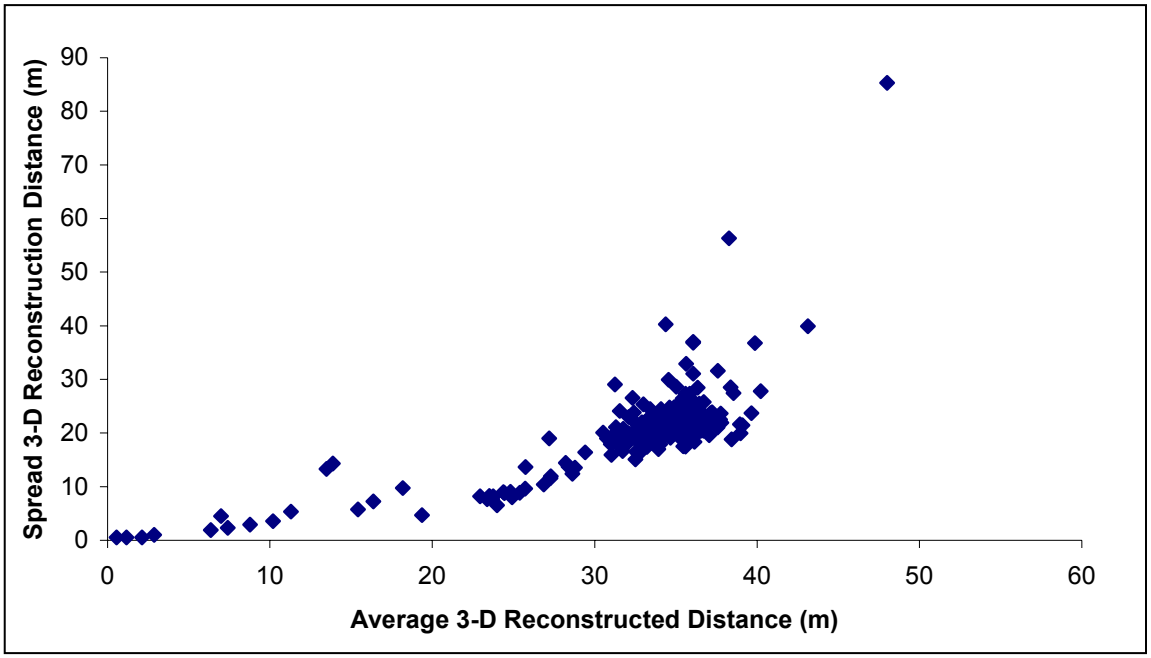

Figure 10. Spread (maximum value - minimum value) versus average distance using calibration parameters from calibration image sets 1 through 5 for the confirmation image pair.

If the variability in the camera parameters is reduced as in those for the calibration image sets 2 through 5, the magnitude of the spread is significantly reduced (see figure 12).

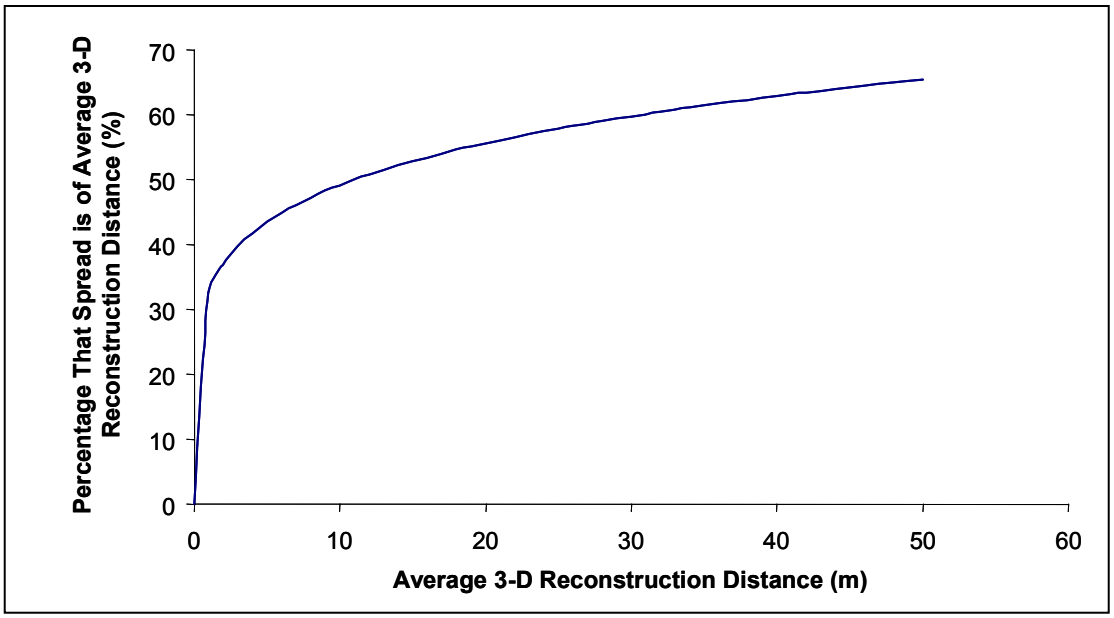

Figure 11. Spread as a percent of average 3-D reconstruction distance for data in figure 10.

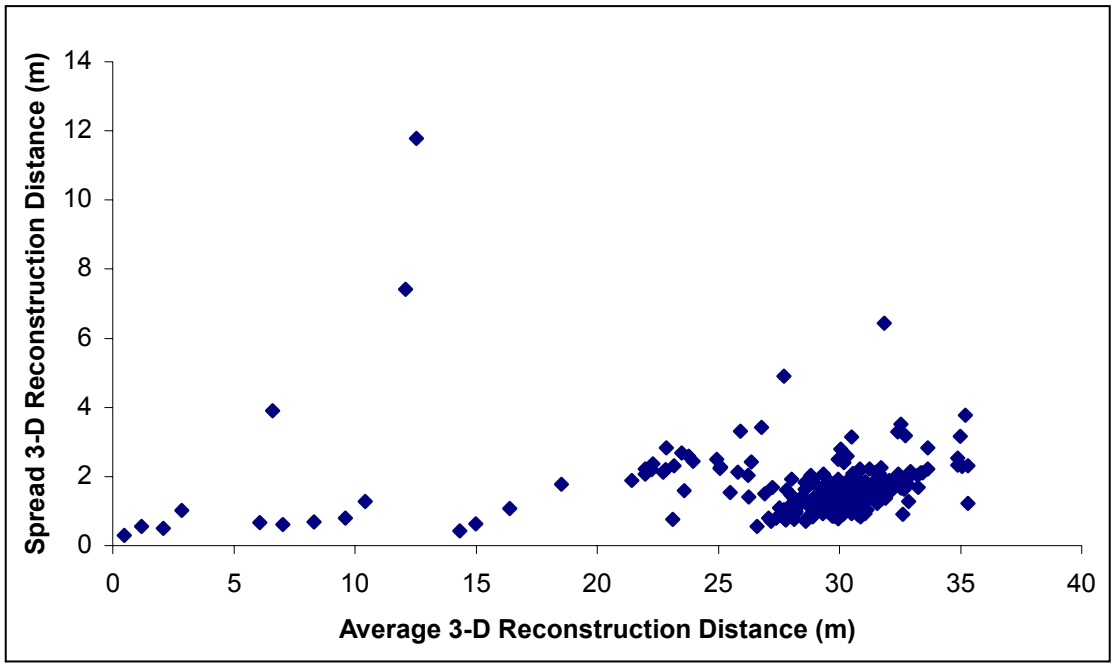

Figure 12. Spread (maximum value - minimum value) versus average distance using calibration parameters from calibration image sets 2 through 5 for the building image pair.

The best fit to the data in figure 12 is $y = 0.5697x^{0.2938}$. With a similar approach as before, the spread as a percentage of average 3-D reconstruction distance for the data is provided in figure 13.

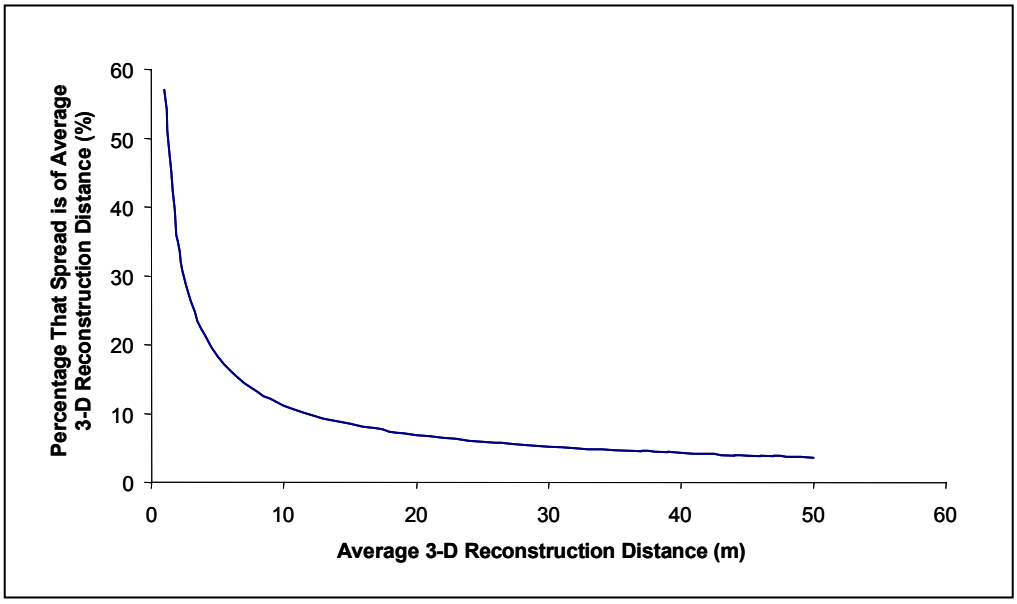

Figure 13. Spread as a percent of average 3-D reconstruction distance for data from figure 12.

To summarize for the building image pair, it appears that if the variability in the camera parameters is approximately the same as that associated with calibration image sets 2 through 5, the resulting distance calculations may exhibit an acceptable degree of variability if less than a 10% spread as a function of average distance is acceptable (see figure 13). The variability associated with calibration image sets 1 through 5 appears to be too large. It is noted that the building image pair is limited to distances of about 50 meters. Generalization of these results to other image pairs, especially image pairs involving greater and a wider spread of distances, is unknown. This question is addressed below for the intersection image pair.

3.2 Results Intersection Image Pair

For the intersection image pair, 919 matched corner points (see figure 14) were identified. In this image pair, the corner points cover a wider range of distances, foreground to down the road, than in the building image pair.

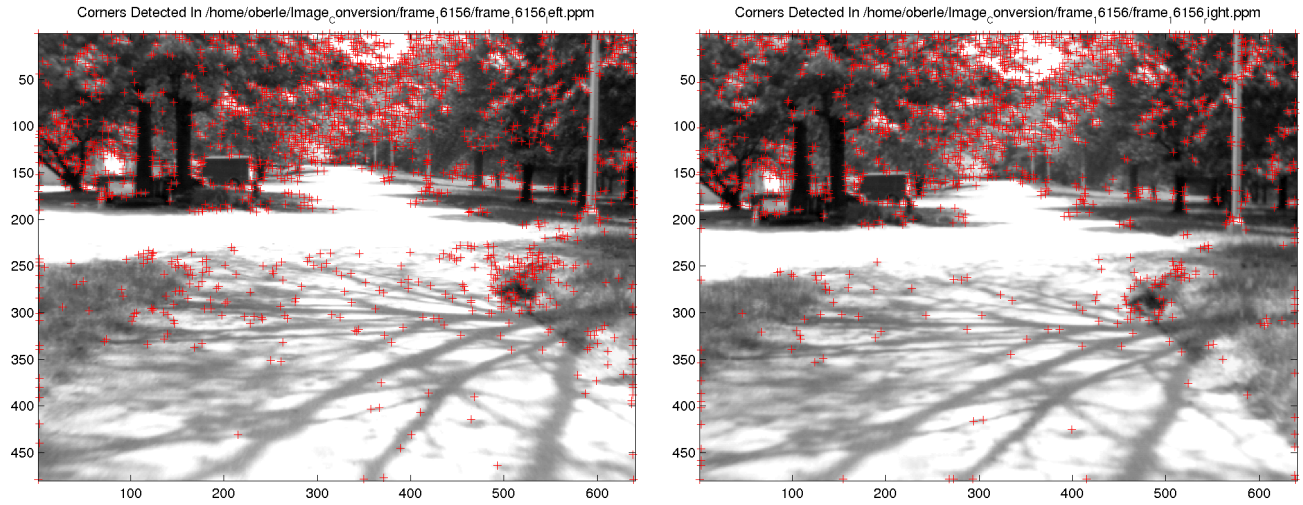

Figure 14. Corner points (+) as determined by the Harris corner detector for the intersection image pair (left image on the left and right image on the right).

Results for the 3-D reconstruction are shown in figure 15. Of the 919 matched points, 507 were considered outliers, based on a negative forward coordinate. Thus, 412 matched points are included in the figure. Although not as clearly discernible as in figure 8, it appears that the 3-D reconstruction distances using calibration image set 1 are consistently larger than the 3-D reconstruction distances for the other calibration image sets. However, since our interest is in the variability of the 3-D reconstruction results, the spread of the distance calculations as a function of average distance needs to be considered. This result for calibration image sets 2 through 5 is shown in figure 16. Similar results for the spread are obtained if the camera parameters for the five calibration image sets are used, and these results are not provided.

Unfortunately, based on figure 16, it appears that the results for the spread in the calculated 3-D reconstruction distances using calibration image sets 2 through 5 show greater variability

than for the building image pair beyond an average distance of 40 meters. Beyond 40 meters' average distance, the spread in general exceeds 25% of the average distance.

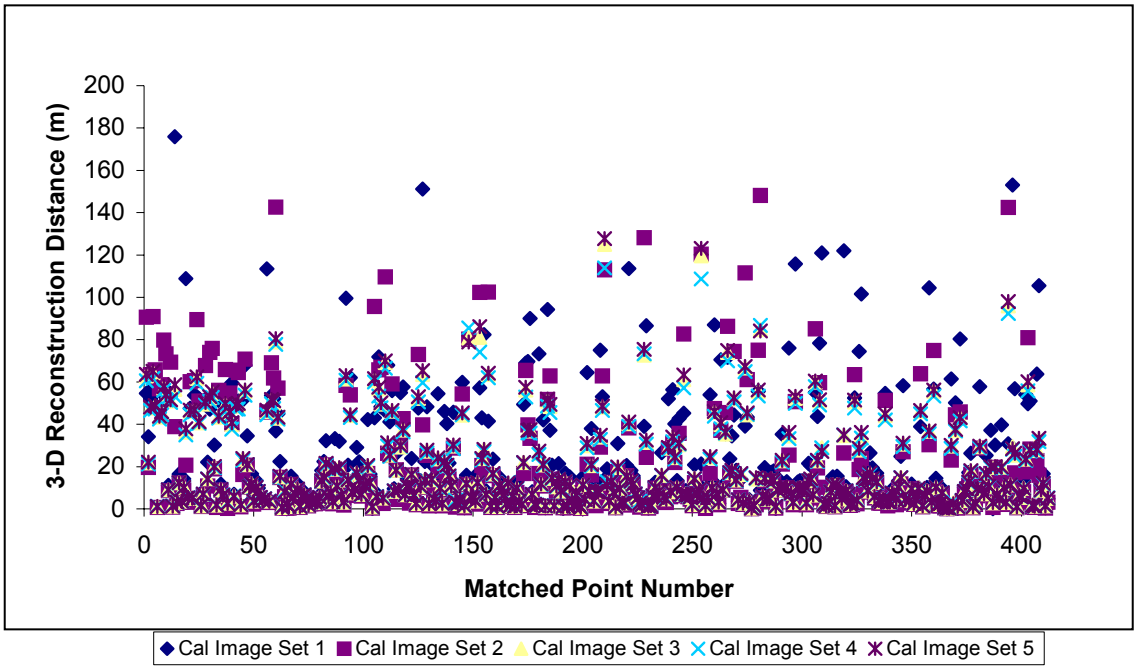

Figure 15. 3-D reconstruction results for intersection image pair using the five different sets of camera parameter.

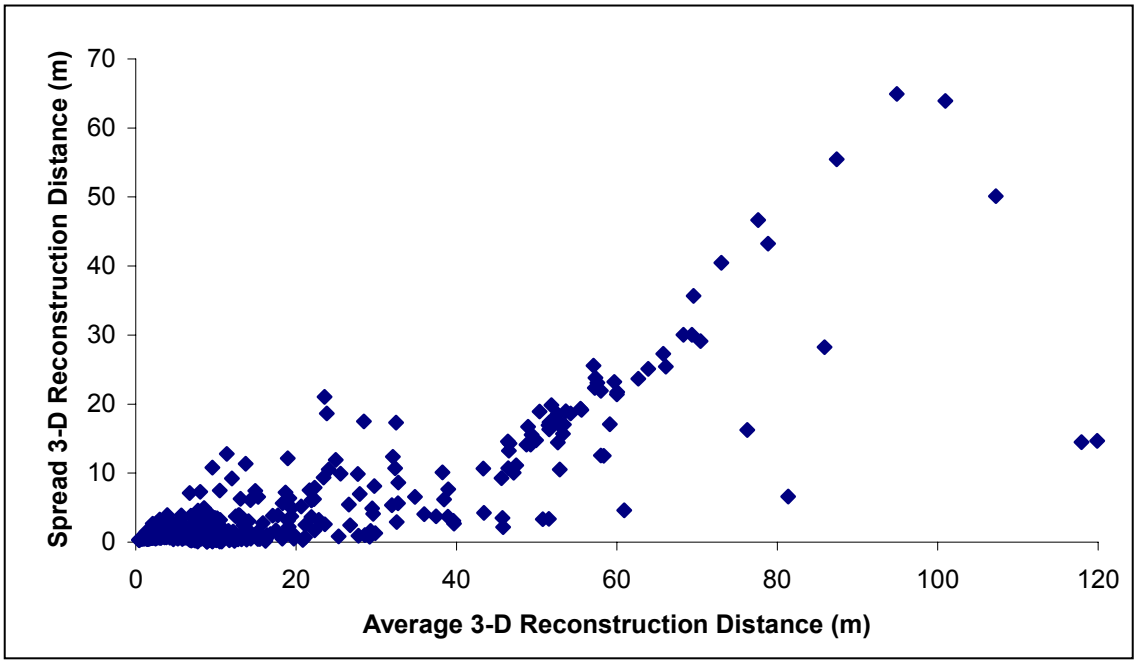

Figure 16. Spread (maximum value - minimum value) versus average distance for calibration image sets 2 through 5 using intersection image pair.

Based on the results of the analysis of the 3-D reconstruction for the building and intersection image pairs, the amount of variability in the camera parameters associated with the five calibration image sets is too large to provide useful 3-D reconstruction distance estimations at most distances. Although the variability associated with calibration image sets 2 through 5 provided potentially acceptable results for the building image pair, beyond 40 meters, the variability (spread) in the distance calculation increases to over 25% of the average distance[10] for the intersection image pair. Thus, for scenes involving distances of more than 40 meters, acceptable 3-D reconstruction may require less variability in the camera parameters than presented in tables 2, 3, and 5 through 8.

In the remainder of this section, we attempt to quantify the amount of variability in the camera parameters that may produce acceptable 3-D reconstruction estimates. For this report, we consider a 10% difference or variability in the 3-D reconstruction distance estimate as acceptable. We make the assumption that if the variability in the camera parameters does not provide reasonable 3-D reconstruction results for distances of <40 meters, then the 3-D reconstruction beyond 40 meters will also be unacceptable. We further assume that the 3-D reconstruction results from variations in the camera parameters are independent, i.e., the 3-D reconstruction result when two parameters are varied is the sum of the 3-D reconstruction result when each parameter is varied individually. Thus, it is sufficient to vary the camera parameters one at a time. Instead of using the camera calibration results for the five calibration image sets, a single set of camera parameters is selected as a baseline. 3-D reconstruction for variations about this baseline are performed and the average percent difference in the resulting distances for all the matching points is computed. Since it appears that the variability in the camera parameters using the five calibration image sets is too great, the average of the camera parameters from calibration image sets 2 through 5 is chosen as the baseline camera parameters. The variation in the camera parameters is accomplished by the addition of ±2 and ±1 standard deviations to the baseline value. The standard deviations for like camera parameters, e.g., components of the translation vector, used in the variation are based upon the CoV for the calibration image sets 2 through 5. The rotation for the *registration* parameters is represented by roll, yaw, and pitch angles. Thus, there are 14 camera parameters to be considered. However, since the variation in pixel pitch (table 3) is low, this set of *intrinsic* parameters is treated as constants. This leaves a total of 10 camera parameters to be varied. Table 14 provides the baseline values for the camera parameters, the fixed CoV used for each set of camera parameters, and the actual standard deviation for the calculations. The field image pair is used for the study.

[10]It is important to note that this observation is for this image pair and the specific camera parameter variation used in the 3-D reconstruction.

Table 14. Camera parameters used in the variation.

Camera Parameter	Baseline Value	Standard Deviation (Baseline * CoV)
Image Center (Pixels)		CoV = 0.5%
Horizontal Left Camera	336.06	1.6803
Vertical Left Camera	247.91	1.23955
Horizontal Right Camera	348.47	1.74235
Vertical Right Camera	265.69	1.32845
Pixel Pitch		NA
Horizontal Left Camera	865.47	NA
Vertical Left Camera	865.1	NA
Horizontal Right Camera	852.54	NA
Vertical Right Camera	851.94	NA
Translation Vector (mm)		CoV = 5%
Horizontal	334.798	16.7399
Vertical	-1.725	0.08625
Forward	-4.154	0.2077
Rotation (milliradians)		CoV = 10%
Roll	-16.23303271	1.62330327
Yaw	-10.68089299	1.068089299
Pitch	4.015925	0.4015925

3.3 Results Field Image Pair

In the field image pair, 364 matched corner points (see figure 17) remained after outliers were eliminated, and these points are used in the calculation of the average percent difference.

We start with variations in the image center; results are shown in figure 18. As stated earlier, only one camera parameter is varied at a time. The first observation from figure 18 is that variations in the horizontal component of the image center result in larger percent differences (almost a factor of 2) than equivalent variations in the vertical component. Equivalent variation means not only the same number of standard deviations but also the same absolute variation. For example, the standard deviation for the horizontal components is approximately 1.7 pixels and for the vertical components, 1.3 pixels (see table 14). Thus, the absolute variation is the same when the number of standard deviations for the vertical component is approximately 1.3 times the standard deviation for the horizontal components (1.7 pixels/1.3 pixels). Not surprisingly, the second observation is that the results for the left and right cameras are approximately reflections of each other about the vertical axis. For example, a +2 standard deviation adjustment in the left camera image center, either horizontal or vertical, yields the same results as a –2 standard deviation adjustment in the corresponding right camera image center. Finally, the more interesting question is how large a variation in the image center in terms of pixels results in acceptable 3-D reconstruction distance estimates (i.e., less than 10% difference). With figure 18, the number of standard deviations from the baseline value for each component can be

determined so that the percent difference in the 3-D reconstruction distance is less than 10%. Multiplying this value by the standard deviation for the component provides the desired pixel value. Details for the image center parameters are provided in table 15.

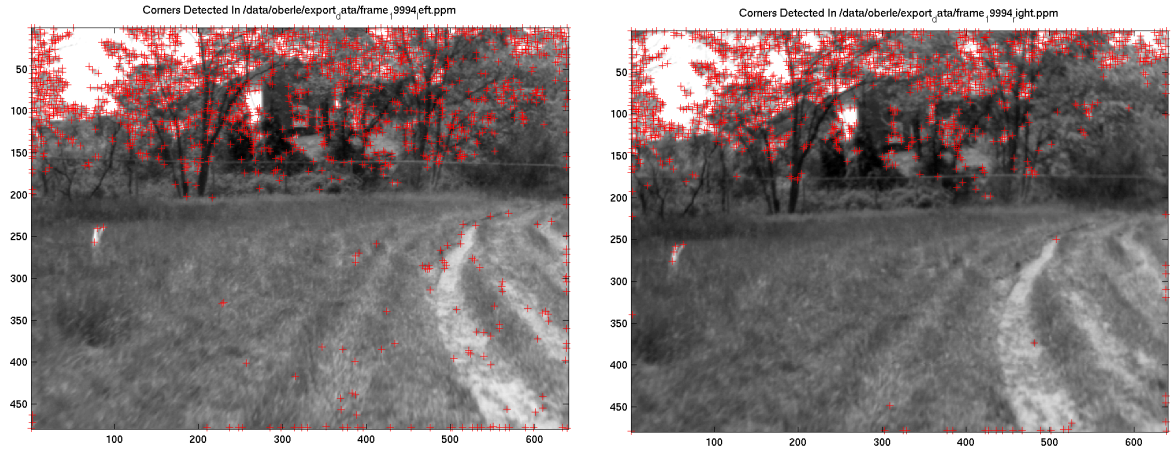

Figure 17. Corner points (+) as determined by the Harris corner detector for the field image pair (left image on the left and right image on the right).

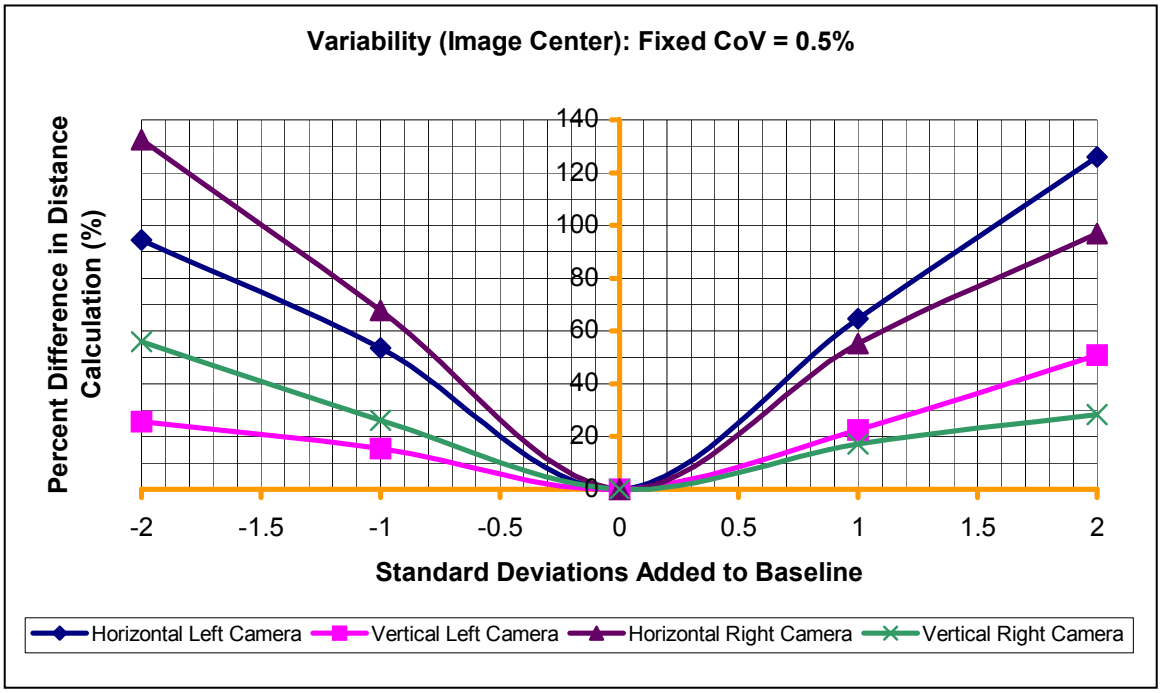

Figure 18. Results for variation in image center, fixed CoV as in table 14.

Table 15. Standard deviations and pixels corresponding to a 10% difference, based on figure 18 and table 14.

Image Center Component	Standard Deviation (pixels) (table 14)	Standard Deviation @ 10% Difference (figure 18)	Pixels @ 10% Difference (Col 2 * Col 3) Maximum Permissible Variation
Horizontal Left Camera	1.6803	0.28	0.470484
Vertical Left Camera	1.23955	0.54	0.669357
Horizontal Right Camera	1.74235	0.28	0.487858
Vertical Right Camera	1.32845	0.50	0.664225

The results in table 15 indicate that the horizontal components of the image center need to be accurate to about 0.5 pixel, and the vertical components need to be accurate to within about 0.7 pixel if the percent difference in the distance calculation is to be less than 10%. Since these results are based on the variation of only one parameter at a time and if our assumption that variations in the parameters produce independent results is valid, then these values for the maximum pixel error in the image center components should represent upper boundaries. The results for the variation in the translation vector components are given in figure 19.

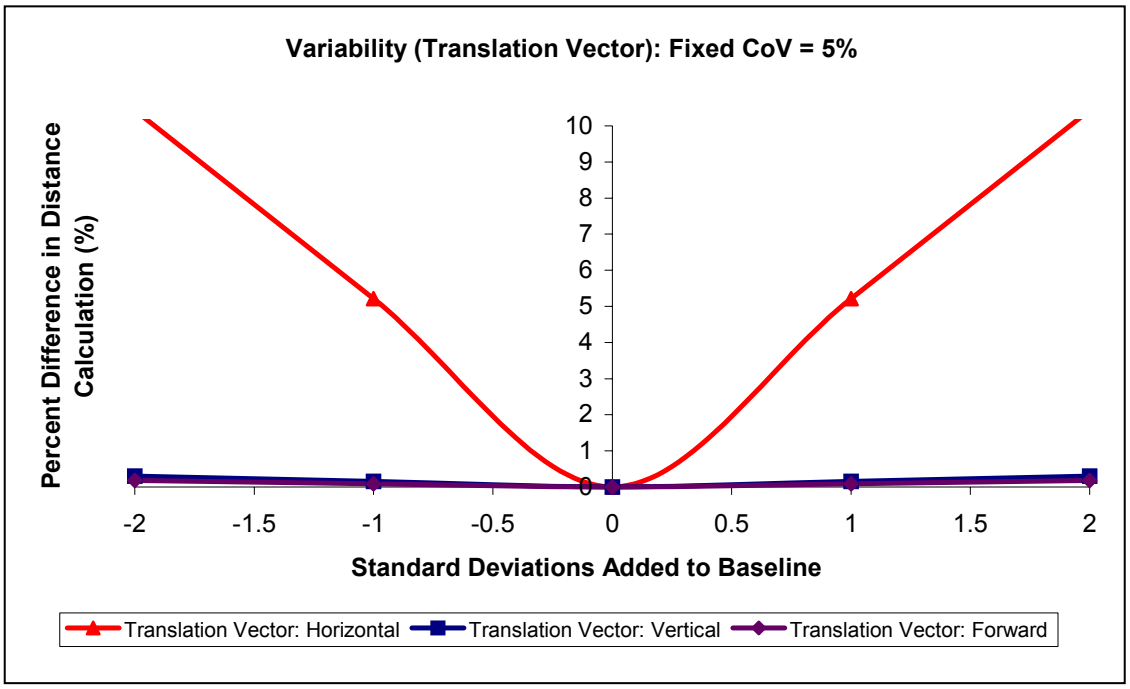

Figure 19. Results of variations in the translation vector, fixed CoV as in table 14.

As indicated in figure 19, variations in the translation vector have a greatly reduced impact on the percent difference in the 3-D reconstruction distance estimates compared to the results in figure 18 for variations in the components of the image centers. The maximum percent error is approximately 10% versus the 120% in figure 18. However, in the case of the translation vector, using a fixed CoV to determine the standard deviations may not be appropriate. From table 5,

the experimentally observed standard deviations for the components are 0.141, 0.114, and 2.343 millimeters, respectively. The corresponding standard deviations used in the calculations from table 14 are 16.7399, 0.08625, and 0.2077 millimeters. Thus, the horizontal component standard deviation is two orders of magnitude high while the forward component standard deviation is an order of magnitude low. Performing the calculation with the experimentally observed standard deviations in table 5 for the components of the translation vector gives the results in figure 20. These results are felt to be more representative of what should be expected in an actual situation. The overall percent differences have decreased and the dominant component in determining the percent difference has changed from the horizontal component (figure 19) to the forward component (figure 20). Fortunately, no matter which set of standard deviations is used, the maximum percent error is below the 10% difference level selected as our acceptable level. Note that for the image center components, the standard deviations in table 2 are approximately the same values as the standard deviations in table 14 used for the calculations depicted in figure 18. Thus, the results and discussion presented earlier for the image center components remain the same even if the experimental standard deviations from table 2 are used.

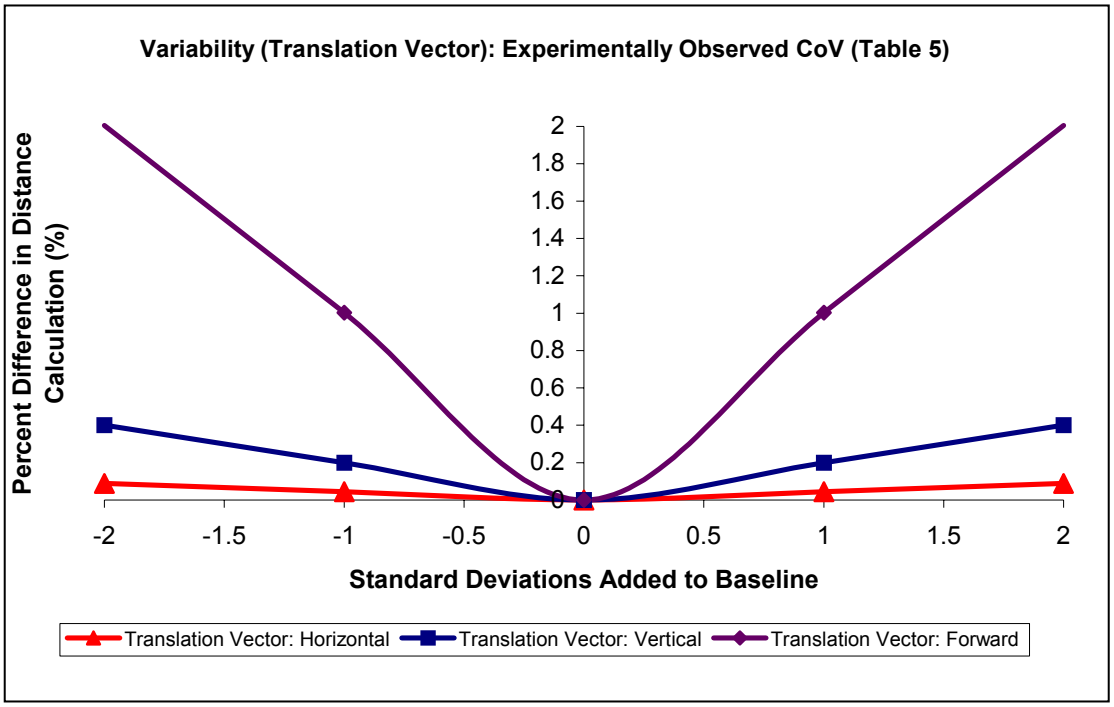

Figure 20. Results of variation in translation vector, experimental standard deviations from table 5.

Finally, the impact of variations in the rotation matrix is presented. Results using the fixed CoV standard deviations from table 14 are shown in figure 21, and the experimentally observed standard deviations from table 11 are shown in figure 22.

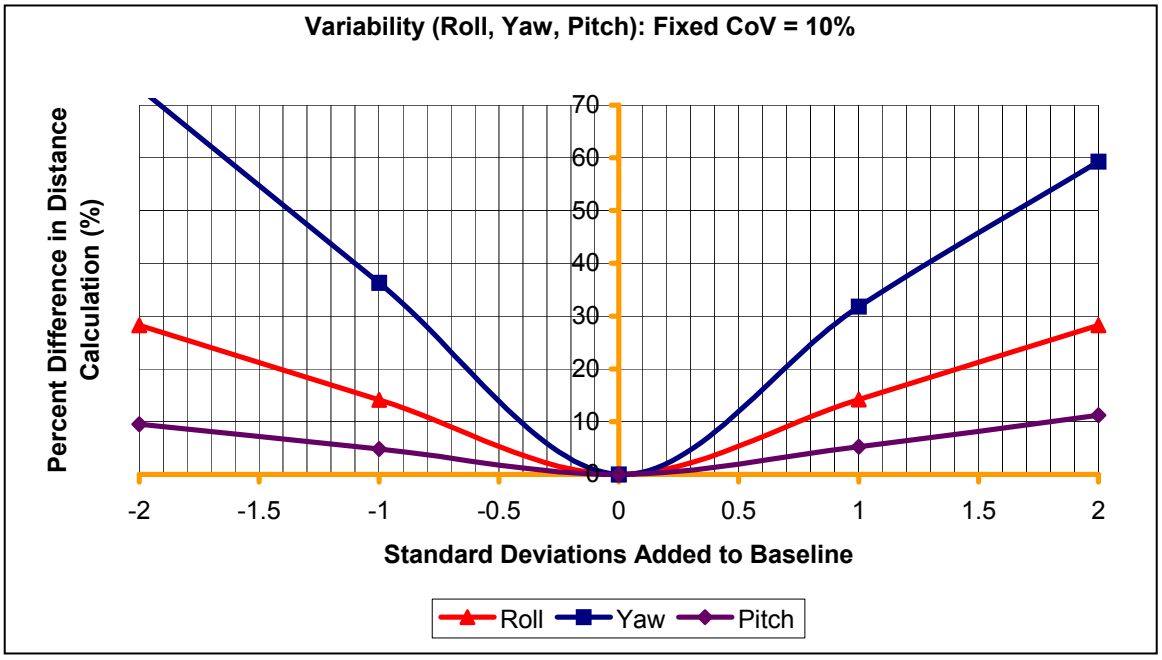

Figure 21. Results of variations in rotation matrix, fixed CoV as in table 14.

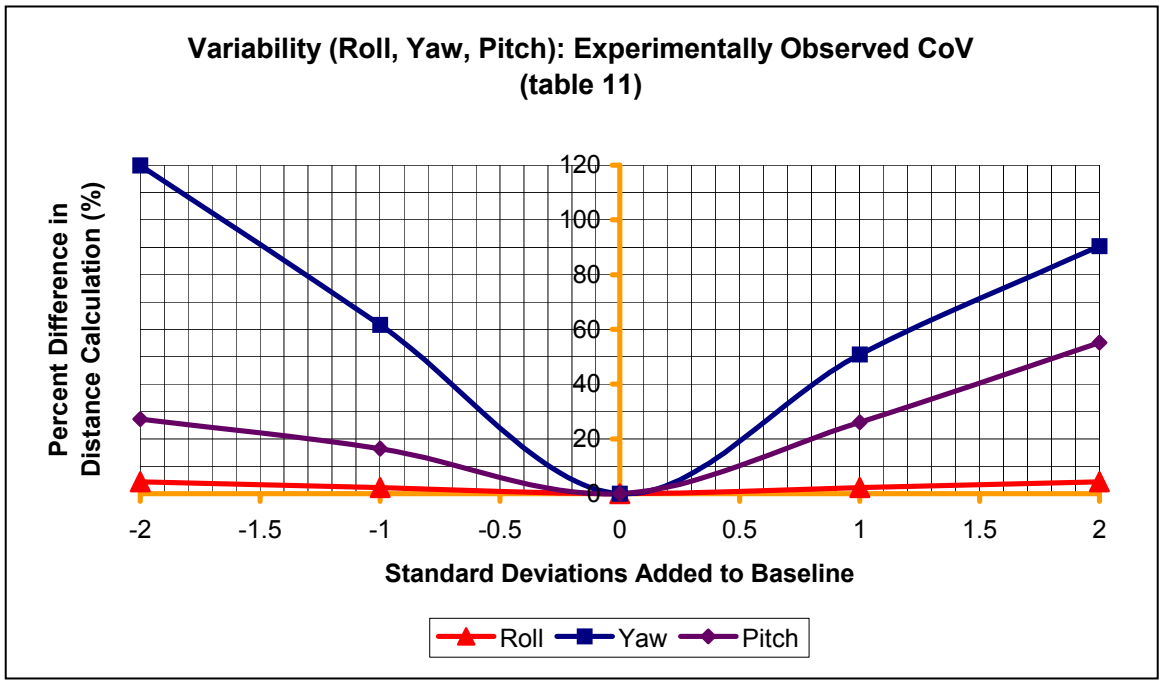

Figure 22. Results of variations in rotation matrix, experimental standard deviations from table 11.

A comparison of the results in figures 21 and 22 indicates that the dominant angle contributing to the percent difference in the 3-D reconstruction distance is the yaw angle. The standard deviation for the yaw angle that produced the results in figure 21 is 1.068089299 milliradians (table 14) compared to a standard deviation of 1.62330327 milliradians (table 14) for the roll

angle. Yet the percent difference associated with variations in the yaw angle is significantly larger compared to the percent difference associated with the roll angle. For figure 22, the standard deviation for the yaw angle is 1.7926 milliradians and for the pitch angle, 1.549 milliradians. However, in figure 22 the percent difference associated with the yaw angle is still substantially higher than the percent difference associated with the roll angle in figure 21 that has about the same standard deviation (1.549 versus 1.62330327 milliradians). The same is true for the yaw and pitch angles in figure 22, i.e., approximately the same standard deviation, yet much higher percent difference associated with the yaw angle.

When an analysis is performed similar to the one performed for the components of the image center, the number of milliradians by which the roll, yaw, and pitch angles can vary in order to achieve at most a 10% difference in the 3-D reconstruction distance estimates yields the results in table 16. These results support the observation that the yaw angle is the dominant angle since the yaw angle has the smallest maximum permissible variation (0.427 milliradian).

Table 16. Standard deviations and milliradians corresponding to a 10% difference, based on figure 21 and table 14.

Rotation Matrix Angle	**Standard Deviation (milliradians) (table 14)**	**Standard Deviation @ 10% Error (figure 21)**	**Milliradians @ 10% Error (Col 2 * Col 3) Maximum Permissible Variation**
Roll	1.62330327	0.75	1.217
Yaw	1.068089299	0.4	0.427
Pitch	0.4015925	1.8	0.723

Several additional calculations are presented in an attempt to place tighter boundaries on the allowable maximum permissible variations in the calibration parameters to maintain the 10% maximum difference when all parameters are varied. For the calculations, we assume that the amount of each parameter's variation has an independent normal distribution. The mean for the parameter is the baseline value in table 14. Details about the standard deviations used in the calculations are described next. Each calculation consists of 1,000,000 iterations. Each iteration starts by our selecting a random number between 0 and 1 for each of the parameters. This random number is interpreted as the area under the normal curve, and the number of standard deviations associated with this area is computed, i.e., inverse of cumulative normal distribution function. The parameter's value for the iteration is set to the baseline value plus the computed number of standard deviations times the standard deviation. Once the values of the ten varying parameters are established, the average percent absolute difference in the 3-D reconstruction distance estimates versus the 3-D reconstruction distance estimates using the baseline values of the parameters for the matched points is computed. The output of the calculation is the frequency distribution of the percent differences.

Earlier, we estimated the maximum pixel variation permissible in the image center components (table 15) and the maximum milliradian variation permissible in the roll, yaw, and pitch angles for the rotation matrix (table 16) necessary to limit the percent difference in 3-D reconstruction distance estimates to a maximum of 10%. For the first calculation, the standard deviations for the parameters are set to one-third these maximum permissible variations. A value of one-third the maximum permissible variation is used since for the normal distribution, 99.7% of the observations will be within ±3 standard deviations of the mean. Since variations in the translation vector contributed little to the percent difference (see figure 20), the experimental standard deviations (see table 5) for the translation vector components are used. Results are presented in figure 23.

As can be observed in figure 23, approximately 53% of the iterations resulted in a percent difference of less than 10% instead of the expected 99.7% if there is no compounding effect because of multiple parameters being varied. In figure 24, the calculation is repeated with the standard deviations reduced by a factor of 2, i.e., one-sixth the maximum permissible variation for the parameters. As expected, the fraction of iterations with an average percent absolute difference below 10% has increased to 90%. Based on these results, it appears that using one-half the values of the maximum permissible variations in the parameters given in tables 15 and 16 for the image center and rotation matrix components should provide an acceptable probability that most of the 3-D reconstructed points have less than a 10% difference in the distance estimate.

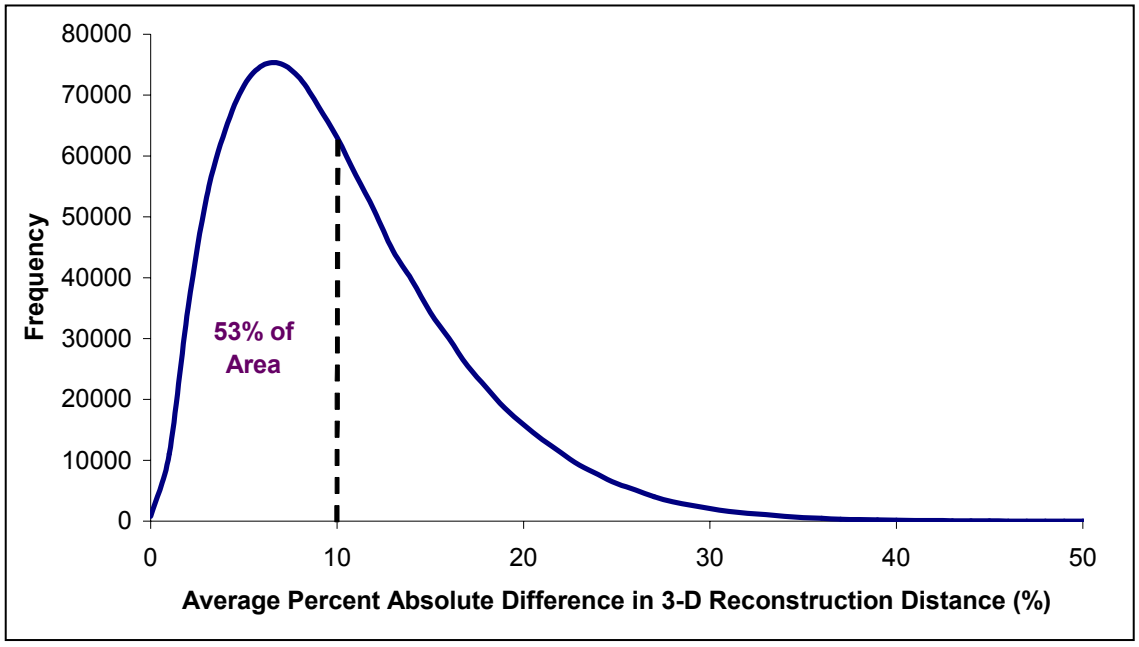

Figure 23. Frequency distribution of absolute percent difference in 3-D reconstruction distance using one-third the maximum permissible variation for parameters as the standard deviation.

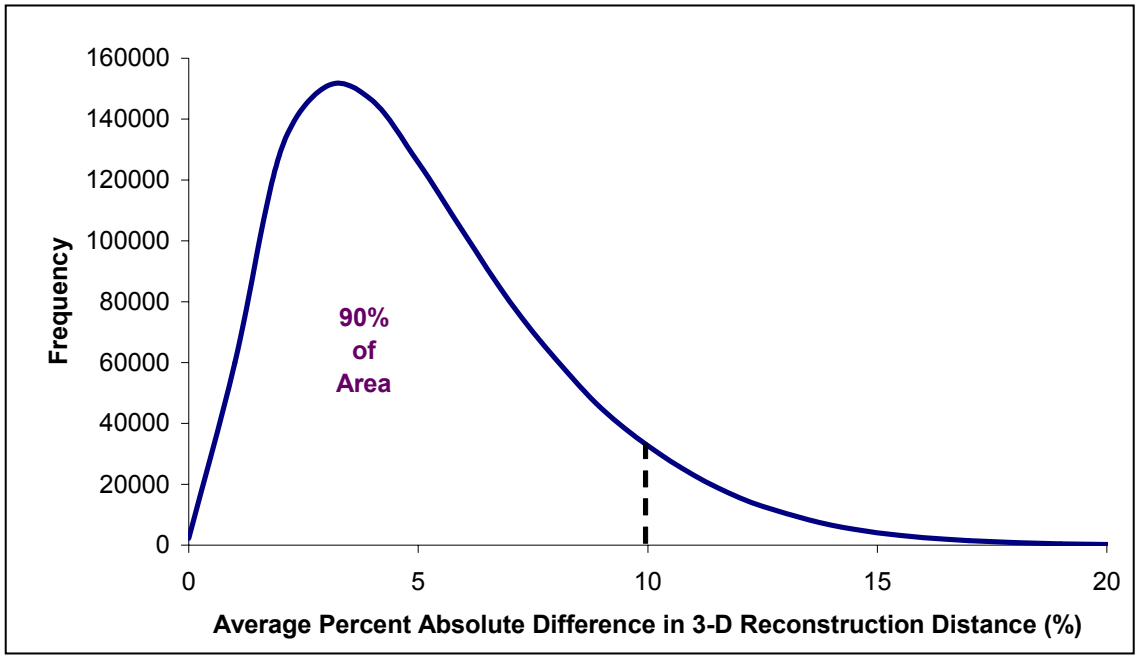

Figure 24. Frequency distribution of absolute percent difference in 3-D reconstruction distance using one-sixth the maximum permissible for parameters as the standard deviation.

Of the three image pairs and five calibrations, the combination most consistent with the limited amount of ground truth is the building image pair and the calibration results using calibration image set 1. To provide a degree of confidence in our belief about the amount of acceptable variability in the calibration parameters, calculations similar to those performed for the field image pair are repeated with this combination of image and calibration values. Specifically, the baseline values for the parameters are the values for calibration image set 1 from tables 2, 3, 5, and 11. The standard deviation values for the parameters in figure 25 are one-third the maximum permissible variations as given in tables 15 and 16 (used for figure 23) and in figure 26, one-sixth the maximum permissible variations as given in tables 15 and 16 (used for figure 24).

In figure 25, 97% of the iterations have an average percent absolute difference of less than 10%. These results are an improvement over the similar calculation for the field image pair in which only 53% of the iterations had an average percent absolute difference of less than 10% (see figure 23). Similar improvements compared to figure 24 are obtained for the final calculation in which one-sixth the maximum permissible variation is used for the standard deviation (see figure 26). In this case, 99.998% of the iterations have an average percent absolute error of less than 10%. Overall, these results are consistent with our earlier comments but probably justify weakening our earlier boundaries from one-half the maximum permissible variation in tables 15 and 16 to the values in the tables.

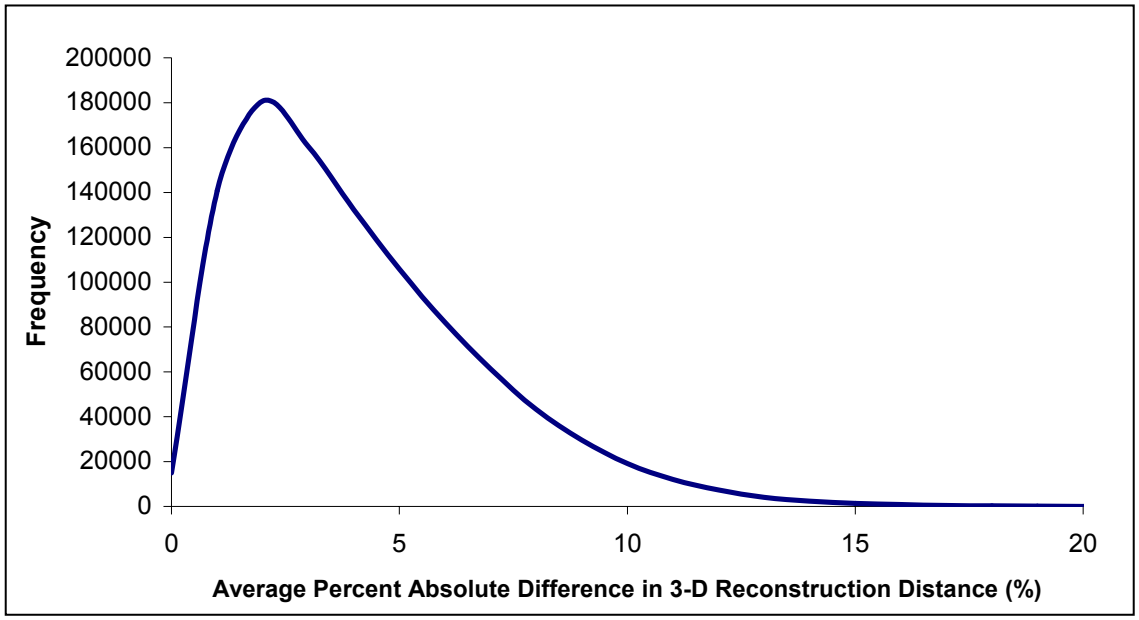

Figure 25. Frequency distribution of absolute percent difference in 3-D reconstruction distance for building image pair with baseline calibration parameters from calibration image set 1 and standard deviations equal to one-third the maximum permissible parameter variation in tables 15 and 16.

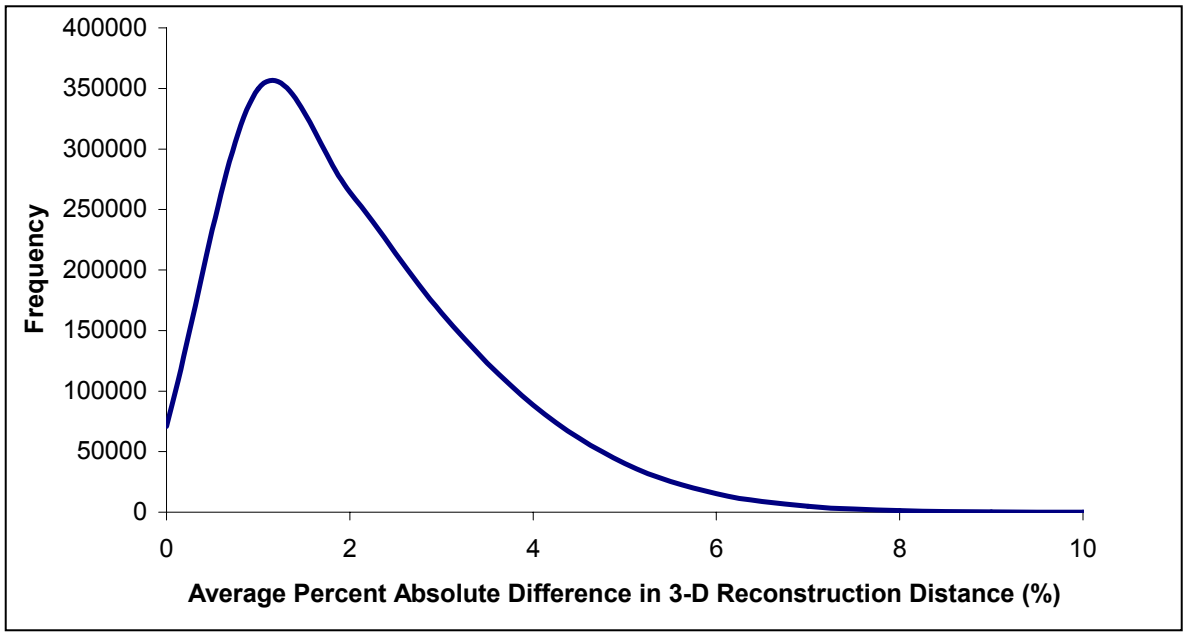

Figure 26. Frequency distribution of absolute percent difference in 3-D reconstruction distance for building image pair with baseline calibration parameters from calibration image set 1 and standard deviations equal to one-sixth the maximum permissible parameter variation in tables 15 and 16.

4. Area-Based 3-D Reconstruction

In the feature-based 3-D reconstruction used in the last section, the original stereo image pair is searched for matching points in the left and right images. For area-based 3-D reconstruction, the stereo image pair is first warped so that matching points are situated on the same horizontal scan or image line. This process is termed "image rectification". Once the images are rectified, matching points are identified through one of several approaches. These approaches all in one form or another essentially search the images to find regions (generally, an n x n neighborhood of pixels) in the left and right images that minimize some cost function. A typical cost function is the sum of the squared differences of the pixel intensities for the pixels in the neighborhood. The fact that the images are rectified implies that for a given point in one image, its matching point must be situated on the same horizontal line in the other image. This reduces the search from two dimensions to one dimension, greatly reducing the search time but at the cost of not finding better or more accurate matches for points on different horizontal lines. Generally, each point in both the left and right images is matched to a point in the other image that minimizes the cost function. Thus, this approach is also termed dense stereo matching.

Rectifying the images requires the use of the *intrinsic* and *registration* parameters. Thus, different sets of calibration parameters may produce different pairs of rectified images. The rectification process subsequently generates a new set of *intrinsic* and *registration* parameters that must be used in the 3-D reconstruction for the matched points from the rectified images. This new set of parameters has properties that are required by the rectification process. First, the left and right image pixel pitch values are all equal. Next, the rotation matrix is the identity matrix. Finally, the vertical and forward components of the translation vector are zero. The resulting rectified image pairs for the building image pair with the calibration parameters associated with calibration image sets 1 through 5 are shown in figures 27 through 31. The rectification of the images is performed with software provided by Haas[11] that is based on the work of Bouguet (2003).

As mentioned earlier, the rectification process produces new *intrinsic* and *registration* parameters associated with the images. These parameters for the rectified images in figures 27 through 31 are provided in tables 17 through 19. In each case, the rotation matrix is the identity matrix.

[11] Private communication, Gary Haas, ARL, APG, Maryland, November 2003.

Figure 27. Rectified image pair for the building image pair with calibration image set 1 parameters.

Figure 28. Rectified image pair for the building image pair with calibration image set 2 parameters.

Figure 29. Rectified image pair for the building image pair with calibration image set 3 parameters.

Figure 30. Rectified image pair for the building image pair with calibration image set 4 parameters.

Figure 31. Rectified image pair for the building image pair with calibration image set 5 parameters.

Table 17. Image center location for rectified images.

Image Center Location (pixels)				
	Left Camera		Right Camera	
Image Set Label/Figure	**Horizontal (x)**	**Vertical (y)**	**Horizontal (x)**	**Vertical (y)**
1 / 27	334.69	257.44	344.19	257.44
2 / 28	338.10	256.46	346.65	256.46
3 / 29	335.05	255.61	347.92	255.61
4 / 30	337.15	258.12	350.16	258.12
5 / 31	335.25	257.62	349.16	257.62
Average	336.05/336.39	257.05/256.95	347.62/348.47	257.05/256.95
SD	1.49/1.48	1.01/1.13	2.32/1.52	1.01/1.13
Range	3.41/3.05	1.66/1.66	5.97/3.51	2.51/2.51
CoV	0.44%/0.44%	0.39%/0.44%	0.67%/0.44%	0.39%/0.44%

Table 18. Pixel pitch for rectified images.

Pixel Pitch (focal length/pixel length)				
	Left Camera		Right Camera	
Image Set Label/Figure	Horizontal (f/sx)	Vertical (f/sy)	Horizontal (f/sx)	Vertical (f/sy)
1 / 27	852	852	852	852
2 / 28	852	852	852	852
3 / 29	853	853	853	853
4 / 30	852	852	852	852
5 /31	852	852	852	852
Average	852.2/852.25	852.2/852.25	852.2/852.25	852.2/852.25
SD	0.45/0.5	0.45/0.5	0.45/0.5	0.45/0.5
Range	1.0/1.0	1.0/1.0	1.0/1.0	1.0/1.0
CoV	0.05%/0.06%	0.05%/0.06%	0.05%/0.06%	0.05%/0.06%

Table 19. Translation vector for rectified images.

Translation Vector (mm)			
Image Set Label/Figure	Horizontal (x)	Vertical (+y down)	Forward (z)
1 / 27	334.559	0	0
2 / 28	334.656	0	0
3 / 29	331.587	0	0
4 / 30	334.686	0	0
5 / 31	334.766	0	0
Average	334.051/333.924	0 / 0	0 / 0
SD	1.379/1.559	0 / 0	0 / 0
Range	3.179/3.179	0 / 0	0 / 0
CoV	0.413%/0.467%	0 / 0	0 / 0

A comparison of tables 17 and 18 with the corresponding tables 2 and 3 shows that for the image center and pixel pitch parameters, there is little change in the variability for the parameters after rectification compared to before the rectification. However, the results shown in tables 5 and 19 indicate a substantial change in the variability of the translation vector. Although the horizontal component variability increased, it is still low, whereas the vertical and forward component variability is zero as required by the rectification. Thus, even though the rotation matrix is the identity matrix in all cases, based on the results on the previous section, we would expect the same level of variability in the 3-D reconstruction since the image center variability (shown to be a substantial contributor to the variability) is approximately the same as in the original parameters, i.e., before rectification. Unfortunately, the situation is more complicated for the area-based 3-D reconstuction. The results from the previous section were predicated on the assumption that we had truly matched points. That assumption is not necessarily valid for the area-based 3-D reconstruction because determining matched points is based on the assumption that the matched points are on the same horizontal image line. An examination of figures 28 through 31 shows that this assumption is not true. Consider the blue line drawn through the image pairs.

In the previous section, we determined that for the building image pair, the calibration parameters for calibration image set 1 were most likely accurate. This is further confirmed in figure 27 where it appears that matching points are on the same horizontal lines, e.g., see the bottom of the window sills above the blue line. In figures 28, 30, and 31, the right image appears to be shifted upward compared to the left image, while in figure 29, the right image appears to be shifted down. Table 20 summarizes the displacement of the rectified right image relative to the rectified left image. The displacement in terms of the number of horizontal lines is determined through an analysis of hand-selected matching image points.

Table 20. Displacement of right image relative to left image for rectified image pairs.

Image Set Label/Figure	Direction Right Image Displaced (relative to left image)	Number of Horizontal Lines Displaced
1 / 27	NA	0
2 / 28	Up	13
3 / 29	Down	5
4 / 30	Up	5
5 / 31	Up	4

As described before, once the images are rectified, matching points are determined under the restriction that the matched points will be situated on the same horizontal scan line. Thus, if the rectification process displaces the images relative to each other, as illustrated in figures 28 through 31, then at best, the matched points as measured in pixels will have an error in the vertical component equal to the number of horizontal lines by which the images are displaced. In the 3-D reconstruction calculation, this equivalent to an error in the vertical image center for either the left or right camera. Referring to table 14, the average standard deviation in the vertical image center is 1.284 pixels ([1.23955 + 1.32845]/2). At two standard deviations (2.568 pixels), the percent difference in the 3-D reconstruction distance calculation is approximately 50% (see figure 18). Therefore, if the horizontal displacement attributable to the rectification is about 2.5 scan lines, a significant variation in the 3-D reconstruction distance is to be expected. Based on the results in table 15, the horizontal image displacement resulting from the rectification process needs to be less than one scan line if the difference in the 3-D reconstruction distance is to be less than 10%.

5. Distance Resolution

As illustrated in the previous two sections, the ability to produce accurate 3-D reconstruction distances depends on the accuracy of the calibration of the stereo camera pair. However, even with precise camera calibration, the use of stereopsis is impacted by the well-known problem associated with distance or range resolution.

Distance resolution is defined as the minimum distance that can be distinguished by the stereo system. The relation between distance resolution, $\Delta dist$, and distance from the stereo camera pair, $dist$, (Videre, 2001, Trucco and Verri, 1998) is

$$\Delta dist = \frac{dist^2}{b * f} * error_{pixelmatch}, \quad (1)$$

in which b is the distance between the camera centers[12] and f is the focal length of the cameras. The relation in equation 1 represents the minimal distance resolution obtained under the assumption that the rays used in the triangulation intersect.

For the camera pair used in our experiments, the distance between the camera centers is approximately 334.794 mm (average of the magnitude of the translation vectors) and the average focal length is 8.588 mm. We obtain the focal length from the pixel pitch by assuming that $s_x = s_y = 0.01$ millimeter as specified by the camera manufacturer. Assuming that the error in the pixel match is one pixel (0.01 mm), the distance resolution and the percentage that the distance resolution is of the distance from the stereo camera pair used in our experimental work is given in figure 32.

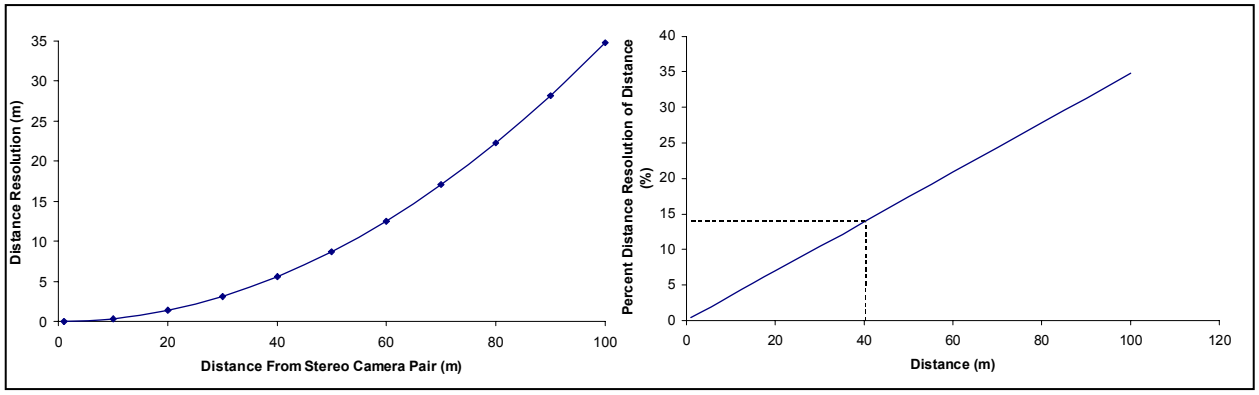

Figure 32. Distance resolution (left) and the percentage that the distance resolution is of the distance (right) from the stereo camera pair used in experimental work.

As can be observed in figure 32, even with perfect camera calibration at 40 meters, the distance resolution is about 5.5 meters or roughly 15% of the distance. At 100 meters, the distance resolution is about 35 meters. Two approaches to improve the distance resolution are to decrease the error in the pixel match, i.e., sub-pixel resolution, and to increase the baseline of the stereo camera pair. Sub-pixel resolution is computationally costly (inhibits real-time application) and depends on the accuracy of the camera parameters. Wide baseline stereo is a more widely used approach and can yield significant reductions in the distance resolution, as illustrated in figure 33 for a variety of camera baselines. The camera focal length (8.588 mm) and error in the pixel match (0.01 mm) are the same as in the previous computation.

[12] Also referred to as the baseline between the cameras.

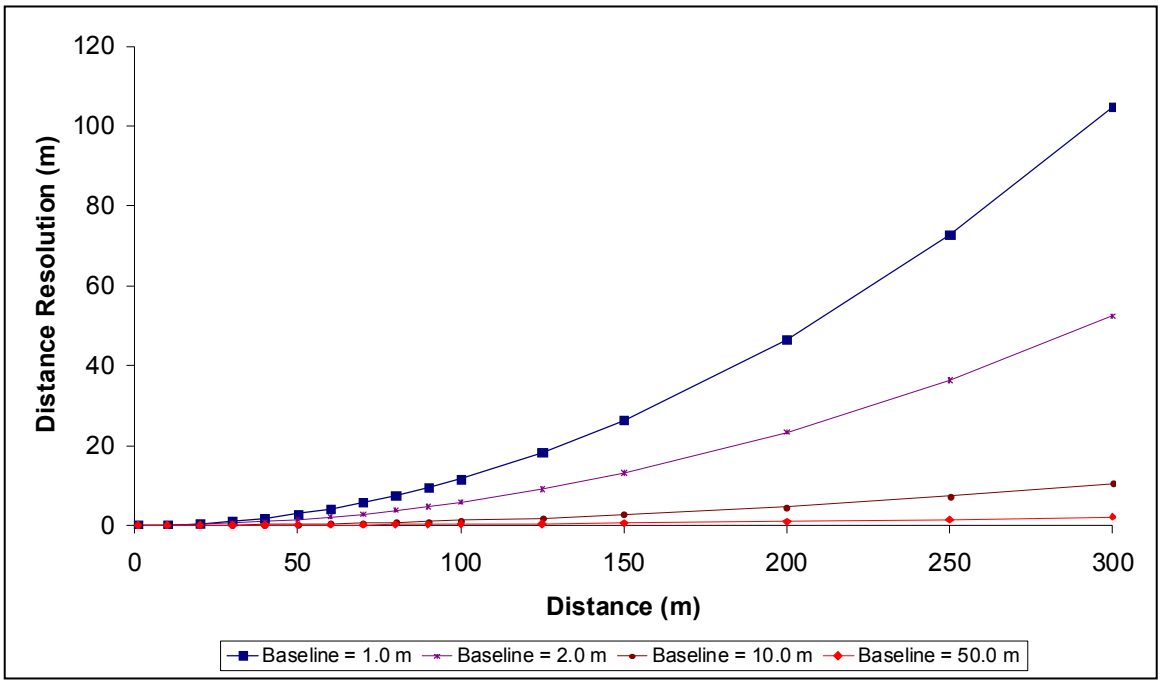

Figure 33. Distance resolution for different camera baselines.

As indicated in figure 33, until the baseline exceeds 10 meters, the distance resolution still exceeds 10% of the distance at distances exceeding ~150 meters. This implies that if distances are to be determined within 10% of the distance from the stereo camera pair, the cameras will have to be mounted on separate vehicles for use with UGVs. However, if the cameras of the stereo pair are mounted on separate vehicles, the camera registration parameters are no longer fixed and have to be determined for each stereo image pair. Thus, we are faced with a difficult situation. High distance resolution requires that the cameras be situated on separate vehicles, but this requires continuous registration[13] between the cameras. Continuous registration will, in all probability, lead to large variability in the *registration* parameters, which could negate the increased resolution attributable to the increased camera baseline. If this problem cannot be circumvented, stereopsis may only be effective for distances as far as 40 meters, even with accurate camera calibration. The decreased distance resolution with increasing distance from the cameras is a result of the triangulation process and cannot be avoided.

[13]Registration between the cameras refers to the process of determining what we have called the camera registration parameters.

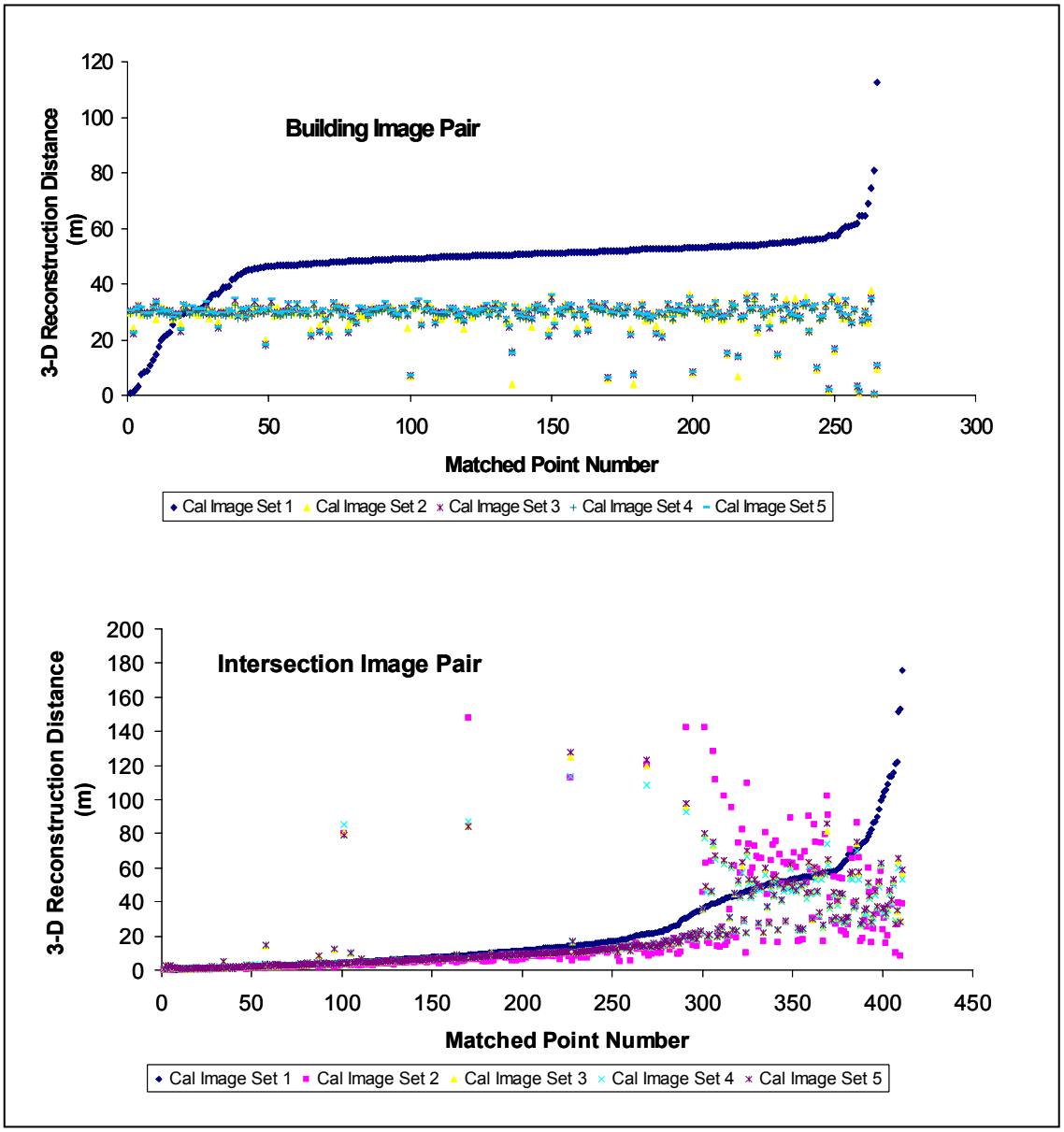

Figure 34. 3-D reconstruction distances, building image pair (top) and intersection image pair (bottom), with the order of the matched points rearranged so that the distance for calibration image set 1 is in increasing order.

6. Summary

This work was motivated by our observation that the camera calibration values calculated for our stereo camera pair tended to vary not only after the stereo cameras had been used to collect data in an outdoor environment but also when repeated calibrations were performed with different

calibration image sets obtained at approximately the same time. These results were observed with the focal lengths of the cameras locked in place. As illustrated in this report, this variability results in poor 3-D reconstruction distance repeatability for both feature-based and area-based 3-D reconstruction. For the specific set of stereo cameras and calibrations used in this report, the variability in the distance calculation was more than 50% of the average calculated distance for distances beyond approximately 40 meters—a value much too large to provide useful information.

The 3-D reconstruction process requires the *intrinsic* and what we have termed *registration* parameters in order to perform the calculation. *Intrinsic* parameters include the camera image center and the pixel pitch for both cameras. The *registration* parameters are the rotation and translation relating the coordiante systems of the left and rigth cameras, or the registration between the two cameras. Based on our results, it appears that the dominant parameters in the degree of variability in the 3-D reconstruction distance calculation are the camera image centers and the rotation matrix of the *registration* parameters. To achieve a percent difference of less than 10% in the calculated 3-D reconstructed distance over a wide range of distances, we estimate that the image centers need to be known to within about one-half a pixel and the yaw, roll, and pitch angles for the rotation to within about 1 milliradian (tables 15 and 16).

However, as discussed in the last section, even perfect knowledge of the calibration parameters does not mean that we can accurately determine distances though stereopsis because of the problem of distance or range resolution. For typical stereo camera pairs, a 10%+ uncertainty in the 3-D reconstruction calculated distance begins at about 35 meters' distance. Wide baseline stereo is a solution to this problem, but the usefulness of this approach may be negated because of the inability to determine the *registration* parameters to the needed accuracy.

Even though the accuracy of the camera calibration may reduce the usefulness of distance calculations via stereopsis, Haas[14] argues that the approach (stereopsis) still provides important information in these cases because the relative position of objects may be correctly determined. He contends that relative location is the information of greatest importance for human vision systems. To investigate the validity of Haas' assertion that relative location is maintained even with variability in the calibration parameters, the distance results for the building and intersection (wider range of distances for matched points) image pairs are analyzed. In figure 34, a scatter plot of the 3-D reconstruction distances for the two image pairs with the five different calibration sets is presented with the order of the matched points rearranged so that the distances using calibration image set 1 are in increasing order. An examination of figure 34 provides little evidence that the relative order of objects is maintained with the amount of variability in the calibration parameters associated with the five calibration image sets. Looking at the same data with the increasing order in the distance based on calibration image set 2 (figure 35), however, does indicate that relative location is maintained if the variability in the calibration parameters is

[14]Private communication, Gary Haas, ARL, APB, Maryland, December 2003.

no greater than that associated with calibration image sets 2 through 5. Thus, it appears that maintaining the relative order of objects with 3-D reconstruction also depends on the amount of variability in the calibration parameters.

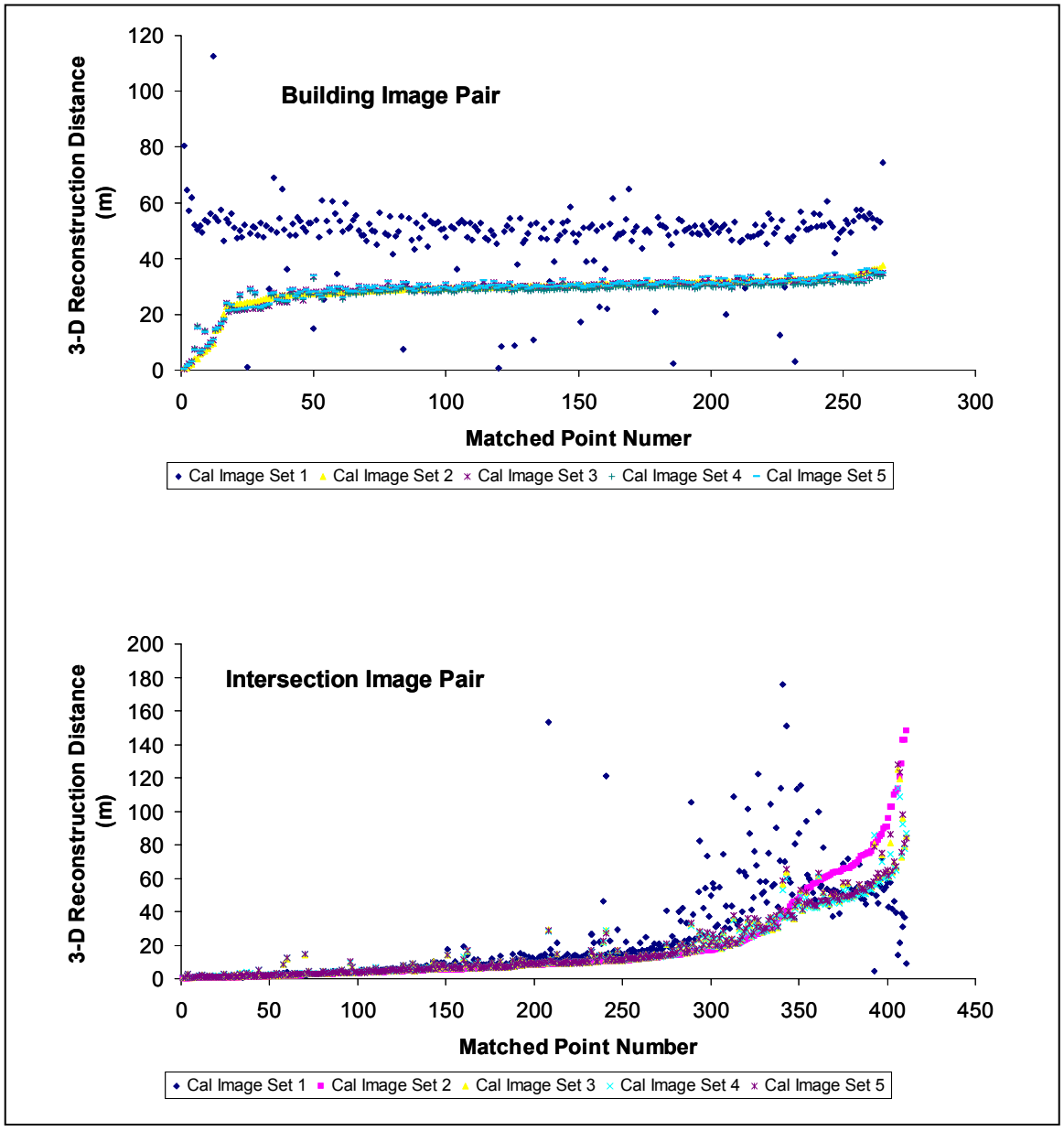

Figure 35. 3-D reconstruction distances, building image pair (top) and intersection image pair (bottom), with the order of the matched points rearranged so that the distance for calibration image set 2 is in increasing order.

Finally, the results in this report are based on a specific set of camera calibrations. The question of whether the variability of this set of camera calibrations is typical must be determined. Toward this end, we suggest that the camera calibration process be performed for several different stereo camera pairs subjected to different operations environments and the results

analyzed to determine the extent of camera calibraton variability. Camera calibration should be performed before and after the stereo camera pair is

1. mounted on a vehicle and
2. used to collect data with the cameras mounted on a moving vehicle traversing outdoor terrain.

Calibration should be performed several times during the day while the stereo camera pair remains in a fixed

1. indoor location and
2. outdoor location.

7. References

Bouguet, J.-Y. *Camera Calibration Toolbox for Matlab*, web site: http://www.vision.caltech.edu/bouguetj/, 2003.

Faugeras, O. *Three-Dimensional Computer Vision: A Geometric Viewpoint*, The MIT Press, Cambridge, Massachusetts, 1993.

Gennery, D.B. *Least-Squares Camera Calibration Including Lens Distortion and Automatic Editing of Calibration Points*, Springer Series in Information Sciences, Vol. 34, *Calibration and Orientation of Cameras in Computer Vision*, Eds. Gruen and Huang, Springer-Verlag, Berlin and Heidelberg, 2001.

Harris, C.G.; Stephens, M. *A combined corner and edge detector.*, Proceeding of the 4th Alvey Vision Conference, pages 147 – 151, Manchester, United Kingdom, 1988.

Jet Propulsion Laboratory, *Cahvor Camera Model Information*, web site: http://robotics.jpl.nasa.gov/people/mwm/chavor.html, California Institute of Technology, Pasadena, California, 2002.

Litwin, T. *CCALDOTS*, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, Written: 13 Mar 1991, Updated: 22 Jun 2000, Copyright (C) 1991, 1992, 1993, 1994, 1996, 1997, 1998, 1999, 2000, California Institute of Technology, All Rights Reserved.

Oberle, W.F.; Haas, G.A. *Three-Dimensional Stereo Reconstruction and Sensor Registration With Application to the Development of a Multi-Sensor Database*; ARL-TR-2878; U.S. Army Research Laboratory: Aberdeen Proving Ground, MD, December 2002.

Oberle, W.F. *Stereo Camera Re-calibration and the Impact of Pixel Location Uncertainty*; ARL-TR-2979; U.S. Army Research Laboratory: Aberdeen Proving Ground, MD, May 2003.

The MathWorks, Inc. MATLAB, Version 6.1.0.450, Release 12.1, web site: http://www.mathworks.com, 2001.

Torr, P.H.S. *A Structure and Motion Toolkit in MATLAB, "Interactive Adventures in S and M,"* Technical Report MSR-TR-2002-56, Microsoft Research, Cambridge, United Kingdom, web site: http://research.microsoft.com/~philtorr/, June 2002.

Trucco, E.; Verri, A. *Introductory Techniques for 3-D Computer Vision*, Prentice Hall, Inc., Upper Saddle River, New Jersey, 1998.

Videre, *STH-DCAM Stereo Head User's Manual*, Updated – Revision 2, Videre Design, Menlo Park, California, 2001.

Yakimovsky, Y.; Cunningham, R.T. *A System for Extracting Three-Dimensional Measurements from a Stereo Pair of TV Cameras*, Computer Graphics and Image Processing 7, pages 195-210, 1978.

NO. OF
COPIES ORGANIZATION

* ADMINISTRATOR
DEFENSE TECHNICAL INFO CTR
ATTN DTIC OCA
8725 JOHN J KINGMAN RD STE 0944
FT BELVOIR VA 22060-6218
*pdf file only

1 DIRECTOR
US ARMY RSCH LABORATORY
ATTN AMSRD ARL CI IS R REC MGMT
2800 POWDER MILL RD
ADELPHI MD 20783-1197

1 DIRECTOR
US ARMY RSCH LABORATORY
ATTN AMSRD ARL CI OK TECH LIB
2800 POWDER MILL RD
ADELPHI MD 20783-1197

1 DIRECTOR
US ARMY RSCH LABORATORY
ATTN AMSRD ARL D D SMITH
2800 POWDER MILL RD
ADELPHI MD 20783-1197

1 DIRECTOR
US ARMY RSCH LABORATORY
ATTN AMSRD ARL SE SE N NASRABADI
2800 POWDER MILL RD
ADELPHI MD 20783-1197

1 NATL INST OF STDS & TECHNOLOGY
ATTN DR M SHNEIER
BLDG 200
GAITHERSBURG MD 20899

1 UNIV OF MARYLAND
INST FOR ADV COMPUTER STUDIES
ATTN DR L DAVIS
COLLEGE PARK MD 20742-3251

ABERDEEN PROVING GROUND

2 DIRECTOR
US ARMY RSCH LABORATORY
ATTN AMSRD ARL CI OK (TECH LIB)
BLDG 305 APG AA

2 DIRECTOR
US ARMY RSCH LABORATORY
ATTN AMSRD ARL WM J SMITH
T ROSENBERGER
BLDG 4600

NO. OF
COPIES ORGANIZATION

2 DIRECTOR
US ARMY RSCH LABORATORY
ATTN AMSRD ARL WM B A HORST
W CIEPIELLA
BLDG 4600

12 DIRECTOR
US ARMY RSCH LABORATORY
ATTN AMSRD ARL WM BF H EDGE
P FAZIO M FIELDS G HAAS
T HAUG W OBERLE (4 CYS)
R PEARSON R VON WAHLDE
S WILKERSON
BLDG 390

2 DIRECTOR
US ARMY RSCH LABORATORY
ATTN AMSRD ARL WM RP C SHOEMAKER
J BORNSTEIN
BLDG 1121